

QY 982 GGTATTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTCTGCTGGGGGTTATCTTATAGTGCCT 1041
Db 1 GGTATTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTCTGCTGGGGGTTATCTTATAGTGCCT 60
QY 1042 CTCATGAATCGGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGTGTGGCAGCTGGCGTTGGG 1101
Db 61 CTCATGAATCGGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGTGTGGCAGCTGGCGTTGGG 120
QY 1102 ACTTTTCAGTGGTGATGCTTTTTTACACCTTTCTTCCACATTTCTCATGTCATGCAAGTCACACCAT 1161
Db 121 ACTTTTCAGTGGTGATGCTTTTTTACACCTTTCTTCCACATTTCTCATGTCATGCAAGTCACACCAT 180
QY 1162 AGTCATAGCCATGAAGAACACAGCAATGAAATGAAAGAGAGCACTTTTTCAGTCATCTG 1221
Db 181 AGTCATAGCCATGAAGAACACAGCAATGAAATGAAAGAGAGCACTTTTTCAGTCATCTG 240
QY 1222 TCCTTCTCAAAACATAGAAGAAGTGCCTTATTTGATTCACAGTGGAAGGGTCTTAACAGCT 1281
Db 241 TCCTTCTCAAAACATAGAAGAAGTGCCTTATTTGATTCACAGTGGAAGGGTCTTAACAGCT 300
QY 1282 CTAGGAGGCTGTATTTTCATGTTCTTGTGGAACATCTCTCATGTCATGATCAAAACAAATTT 1341
Db 301 CTAGGAGGCTGTATTTTCATGTTCTTGTGGAACATCTCTCATGTCATGATCAAAACAAATTT 360
QY 1342 AAAGATAAGAGAAAGAAATCAGAGAAACCTGAAATGATGATGTGGAGATTAAAG 1401
Db 361 AAAGATAAGAGAAAGAAATCAGAGAAACCTGAAATGATGATGTGGAGATTAAAG 420
QY 1402 AAGCAGTTGTCCAAGTATGATCTCAACTTTCAACAATCAGGAGAAATGATGATGATGAT 1461
Db 421 AAGCAGTTGTCCAAGTATGATCTCAACTTTCAACAATCAGGAGAAATGATGATGATGATGAT 480
QY 1462 GATCGAATGAAGGCTATTTACAGCAGACTCAACAAGAGCCCTCCCACTTTGATTTCTCAG 1521
Db 481 GATCGAATGAAGGCTATTTACAGCAGACTCAACAAGAGCCCTCCCACTTTGATTTCTCAG 540
QY 1522 CAGCTTCAGCTTTGGAAGAGAGAGGTTCATGATAGCTCATGCTCATGCTCATGCTCATGCT 1581
Db 541 CAGCTTCAGCTTTGGAAGAGAGAGGTTCATGATAGCTCATGCTCATGCTCATGCTCATGCT 600
QY 1582 TACAAATGAATATGTACCCAGAGGTGCAAGAATAAATGCCATTCACATTTTCCAGCATACA 1641
Db 601 TACAAATGAATATGTACCCAGAGGTGCAAGAATAAATGCCATTCACATTTTCCAGCATACA 660
QY 1642 CTCGCCCATGACAGATCTCATTCACCCACCATCATGATACCATCATATTTCTTCATCAT 1701
Db 661 CTCGCCCATGACAGATCTCATTCACCCACCATCATGATACCATCATATTTCTTCATCAT 720
QY 1702 CACCAACCAACCAACCAACCACTCTCAGCTCAGCTCAGCAGCGCTACTCTCGGAGAGCTG 1761
Db 721 CACCAACCAACCAACCAACCACTCTCAGCTCAGCTCAGCAGCGCTACTCTCGGAGAGCTG 780
QY 1762 AAAGATCGCGGCTGCCCACTTTGGCTGGATGGTATGATAGGTGATGCGCTGCAAT 1821
Db 781 AAAGATCGCGGCTGCCCACTTTGGCTGGATGGTATGATAGGTGATGCGCTGCAAT 840
QY 1822 TTACAGCATGCGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAT 1881
Db 841 TTACAGCATGCGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAT 900
QY 1882 ACTTCTGTTGCTGTCTCTCATGAGTTGCTCATGATAGGTGATGCGCTGCAAT 1941
Db 901 ACTTCTGTTGCTGTCTCTCATGAGTTGCTCATGATAGGTGATGCGCTGCAAT 960
QY 1942 CTAAGGCTGGCATGACCTTTAAGCAGGCTGTCTTTTATAATGCAATTTGCAAGTGTG 2001
Db 961 CTAAGGCTGGCATGACCTTTAAGCAGGCTGTCTTTTATAATGCAATTTGCAAGTGTG 1020
QY 2002 GCGTATCTTGGAAATGCAACAGGAATTTTCATTTGCTGATTTATGCTGAAATATTTTCATG 2061
Db 1021 GCGTATCTTGGAAATGCAACAGGAATTTTCATTTGCTGATTTATGCTGAAATATTTTCATG 1080

QY 2062 TGGATATTTGCACTTACTGCTGCTTATTTCATGATGTTGCTGCTGATATGTTGCTGCT 2121
Db 1081 TGGATATTTGCACTTACTGCTGCTTATTTCATGATGTTGCTGCTGATATGTTGCTGCT 1140
QY 2122 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTT 2181
Db 1141 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTT 1200
QY 2182 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATTTGCTTACTTATTTCCATATTTGAACAT 2241
Db 1201 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATTTGCTTACTTATTTCCATATTTGAACAT 1259
QY 2242 AAAATGCTGTTTTCGTATATAAATTTCTAG 2268
Db 1260 AAAATGCTG-TTCGTATATAAATTTCTAG 1285

RESULT 2
US-08-311-023-3
; Sequence 3, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; TITLE OF INVENTION: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-311-023-3

Query Match 55.7%; Score 1263; DB 1; Length 2404;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1285; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 982 GGTATTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTCTGCTGGGGGTTATCTTATAGTGCCT 1041
Db 1 GGTATTTATAGCCATTTCCATCATCAGTTTCCCTGCTCTCTGCTGGGGGTTATCTTATAGTGCCT 60
QY 1042 CTCATGAATCGGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGAGTTTCTTGTGGCAGCTGGCGGTTGGG 1101
Db 61 CTCATGAATCGGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGAGTTTCTTGTGGCAGCTGGCGGTTGGG 120

QY 1102 ACTTTGAGTGGTGATGCTTTTACACATCTTCCACATTTCTCATGCAAGTCAACAGCAT 1161
DB 121 ACTTTGAGTGGTGATGCTTTTACACATCTTCCACATTTCTCATGCAAGTCAACAGCAT 180
QY 1162 AGTCATAGCCATGAAGAACCAAGCAATCGAATGAAAGAGAGACCACTTTTCAATCATCTG 1221
DB 181 AGTCATAGCCATGAAGAACCAAGCAATCGAATGAAAGAGAGACCACTTTTCAATCATCTG 240
QY 1222 TCTTCTCAAAACATAGAAAGTGGCTATTTTGATTTCCACGTGGAGAGTCTAAACAGCT 1281
DB 241 TCTTCTCAAAACATAGAAAGTGGCTATTTTGATTTCCACGTGGAGAGTCTAAACAGCT 300
QY 1282 CTAGAGGCCCTGTATTTTCATGTTTCTTGAACATGCTCTCACATTTGATCAAAACATTT 1341
DB 301 CTAGAGGCCCTGTATTTTCATGTTTCTTGAACATGCTCTCACATTTGATCAAAACATTT 360
QY 1342 AAACATAGAGAAAGAAATCAAGAAACCTCAAAATCATGATGATGATGATGATGATGAT 1401
DB 161 AAACATAGAGAAAGAAATCAAGAAACCTCAAAATCATGATGATGATGATGATGATGATGAT 420
QY 1402 AAGCAGTTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATCAGGAGAAAGTAGATACAGAT 1461
DB 421 AAGCAGTTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATCAGGAGAAAGTAGATACAGAT 480
QY 1462 GATCGAAGTGAAGGCTATTTACGAGCAGACTCAAGAGCCCTCCCACTTTTGATTTCTCAG 1521
DB 481 GATCGAAGTGAAGGCTATTTACGAGCAGACTCAAGAGCCCTCCCACTTTTGATTTCTCAG 540
QY 1522 CAGCTCTGAGCTTTGGAGAGAGAGGTCATGATAGCTCATGCTCATCCACAGAGAGTC 1581
DB 541 CAGCTCTGAGCTTTGGAGAGAGAGGTCATGATAGCTCATGCTCATCCACAGAGAGTC 600
QY 1582 TACAATGAATATGTACCCAGAGGCTGCAAGATAAATGCCATTTCACATTTCCACGATACA 1641
DB 601 TACAATGAATATGTACCCAGAGGCTGCAAGATAAATGCCATTTCACATTTCCACGATACA 660
QY 1642 CTCGAGCAGTCAGAGCTCATTTACCCAGCAGCTATGATGATGATGATGATGATGATGAT 1701
DB 661 CTCGAGCAGTCAGAGCTCATTTACCCAGCAGCTATGATGATGATGATGATGATGATGATGAT 720
QY 1702 CACACACCAAAACCAACCTCTCAGTCAGCAGCAGCGGTACTCTCGGAGAGAGTCG 1761
DB 721 CACACACCAAAACCAACCTCTCAGTCAGCAGCAGCGGTACTCTCGGAGAGAGTCG 780
QY 1762 AAAGATGCTGGCGTGGCCACTTTTGGCGTGGATGATGATGATGATGATGATGATGATGATGAT 1821
DB 781 AAAGATGCTGGCGTGGCCACTTTTGGCGTGGATGATGATGATGATGATGATGATGATGATGAT 840
QY 1822 TTTGAGGATGGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGTGTTAAAGT 1881
DB 841 TTTGAGGATGGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGTGTTAAAGT 900
QY 1882 ACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941
DB 901 ACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 1942 CTAAGGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTATAATGCAATTTGTCAGGATGCTG 2001
DB 961 CTAAGGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTATAATGCAATTTGTCAGGATGCTG 1020
QY 2002 GCGTATCTTGGATGCAACAGAAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2061
DB 1021 GCGTATCTTGGATGCAACAGAAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 2062 TGGATATTTGCACTTACTGCTGCTTATTCATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2121
DB 1081 TGGATATTTGCACTTACTGCTGCTTATTCATCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 2122 GAAATGCTGCAAAATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATGATGATGAT 2181
DB 1141 GAAATGCTGCAAAATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 2182 CACAATGCTGGGATGCTTTTGGGTTTGGAAATVATGTTACTTATTTTCATVATTTCAAAATV 2241

DB 1201 CAGATGCTGGGATGCTTTGGGTTTGGAAATATGTTACTTA-TTCCATATTTGAACAT 1259
QY 2242 AAAATCGTCTTTCGTATAAAATTTCTAG 2268
DB 1260 AAAATCGTCTTTCGTATAAAATTTCTAG 1285

RESULT 3

US-08-990-571-64/c
; Sequence 64, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-990-571-64

Query Match 2.6%; Score 59.8; DB 4; Length 285;
Best Local Similarity 65.2%; Pred. No. 7.3e-08;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGACCCAGCACCATCACTCAGACAGGAGCATCACTCAGACCATGAG 327
DB 192 AACCAAGTTCCACTAGGCCCCAGCAGCTTCACCTAGGCCCCAGCAGCTTCACCTAGGCCCCAGCAG 133
QY 328 COTCACTCAGACCATGAGCATCACTCAGACCCAGGAGCATCACTCTGACCATGATCATCAC 387
DB 132 CTTCACTAGGCCCCAGCAGCTTCACCTAGGCCCCAGCAGCTTCACCTAGGCCCCAGCAGCTTCAC 387
QY 388 TCTCACTAATCAT 402
DB 72 TAGGCCCCAGCAGCTT 58

RESULT 4

US-09-528-784A-64/c
; Sequence 64, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Iodes, Michael J.
; APPLICANT: Houghlon, Raymond L.
; APPLICANT: Sleath, Paul R.


```

: CURRENT APPLICATION NUMBER: US/09/528,784A
:
: CURRENT FILING DATE: 2000-03-17
:
: NUMBER OF SEQ ID NOS: 90
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 63
:
: LENGTH: 356
: TYPE: DNA
:
: ORGANISM: Babesia microti
US-09-528-784A-63

```

Query Match 2.6%; Score 59.8; DB 4; Length 156;
Best Local Similarity 65.2%; Pred. No. 8.4e-08;
Matches 88; Conservative 0; Mismatches 47; Indels

QY	268	ATTCATATATACACCATGACCCAGGACCATCACTACAGACCAGGACGATCACTTCAGACCATGAG	327
UB	210	AACCAAGTTCACCTTAGGCCCCACAGCTTCACTAGGCCACACGATTCACATAGGCCACACGAG	151
QY	328	CCTCACTCAGACCATGAGCATCACTCAGACACGAGCATCACTCTGACCATGATCATCATAC	387
DB	150	CTTCACTAGGCCACACAGCTTCACTAGGCCACACGCTTCACTAGGCCACACAGCTTCAC	91
QY	388	TCTCACCATAATCAT	402
DB	90	TAGGCCACACAGCTT	76

RESULT 11

US-08-990-571-58/c
; Sequence 58, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:

APPLICANT: Reed,

```

1  TITLE OF INVENTION: COMPOUNDS AND
2  NUMBER OF SEQUENCES: 79
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: SEED AND BERRY
5  STREET: 6300 Columbia Center, 7
6  CITY: Seattle
7  STATE: Washington
8  COUNTRY: USA
9  ZIP: 98104
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patentin Release #1.0
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/990,5
17 FILING DATE: 11-DEC-1997

```

```

: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Maki, David J.
: REGISTRATION NUMBER: 31.392
: REFERENCE/DOCKET NUMBER: 210121.426C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206)682-6031
: INFORMATION FOR SEQ ID NO: 58:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 358 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-990-571-58

```

Query Match 2.68; Score 59.8; DB 4; Length 158;

Best local similarity	65.2%	Pred. No. 8.5e-08;	Length 1.06;
Matches	88; Conservative	0; Mismatches	47; Indels
		0;	Gaps

268	ATCCATATACACCATGACCACGACCATCACTCAGACCACGAGCATCATTCAGACCATGAG	127
228	AACCACTTCACTAGGCCCAACCACTTCACTAGGCCCAACCACTTCACTAGGCCCAACCACT	169

Db 228 AACGAGCTTCATGAGCCCCACCAGCTTCACTAGGCCCCACCAGCTTCAC"AGGTCACACAG 169

Qy	328	CGTCACCTACAGACCATGAGCATCACTCAGACCCAGAGCATCACTCTGACCATGATCATCAC	387
Db	168	CTTCACCTAGGCCACCAGCTTCACCTAGGCCACCAGCTTCACCTAGGCCACCAGCTTCAC	109
Qy	388	TCTCACCATAATCAT	402
Db	108	TAGGCCACCAGCTT	94

RESULT 12

US-09-528-784A-58/c
; Sequence 58, Application US/09528784A

; Patent No. 6451315

```

: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Sleath, Paul R.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
: FILE REFERENCE: 210121.426C4
: CURRENT APPLICATION NUMBER: US/09/528,784A
: CURRENT FILING DATE: 2000-03-17

```

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 58

; LENGTH: 35

; TYPE: DNA

ORGANISM: Babesia microti

US-09-528-784A-58

Query Match

Query Match	2.6%	Score 59.8;	DB 4;	Length 358;
Best Local Similarity	65.2%	Pred. No. 8.5e-08;		

Matches	88;	Conservative	0;	Mismatches	47;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Qy	268	ATCCATATACACCATGACCACGACCATCTACTCAGACGACGAGCATCATCTGACCATCATGAG	327
Db	228	AACGAGCTTTCACCTAGGCCACGAGCTTCTACTAGGCCACACGCTTTCATCTAGGCCACCATG	169
Qy	328	CGTCACTGAGACCATGAGCATCATCTCAGACGACGAGCATCTCTGACCATGATCATCATC	387
Db	168	CTTTCATAGGCCACGAGCTTCTACTAGGCCACACGCTTTCATCTAGGCCACCATGCTTCAC	109
Qy	388	TCTCACCATAATCAT	402
Db	108	TAGGCCACGAGCTT	94

RESULT 13

US-08-990-571-66/c
; Sequence 66, Application US/08990571

; Patent No. 6214971

```

: GENERAL INFORMATION:
: APPLICANT: Read, Steven G. et al.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
: NUMBER OF SEQUENCES: 79
: CORRESPONDENCE ADDRESS:
: ADDRESS: SEED AND BERRY
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: Washington
: COUNTRY: USA
: ZIP: 98104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/990,571
: FILING DATE: 11-DEC-1997

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:28:40 ; Search time 5666 seconds
(without alignments)
11649.341 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcaggaagtattctgt.....tgttcgtataaaatttctag 2268

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBankl :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_pl.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2268	100.0	3461	6	AX465588	Sequence
2	2264.8	99.9	2776	6	AX207207	Sequence
3	2222	98.0	2744	9	HSU41060	Homo sapien
4	2198	96.9	3461	6	AX207205	Sequence
5	1616	71.3	3287	10	AB071697	Mus muscu
6	1263	55.7	1310	6	176891	Sequence 1
7	1263	55.7	2404	6	176892	Sequence 3
8	1151	50.7	2236	9	BC008317	Homo sapi
9	789	34.8	160170	2	AP001158	Homo sapi
10	787.4	34.7	221941	9	AC091060	Homo sapi
11	787.4	34.7	224788	2	AP001905	Homo sapi
12	619.6	27.3	899	6	AX465590	Sequence
13	549.4	24.2	551	6	AX017261	Sequence
14	509.4	22.5	1597	6	AX207216	Sequence
15	421.4	18.6	193437	2	AC126667	Rattus no
16	319.2	14.1	337	6	AX393861	Sequence
17	319	14.1	5231	9	AB033091	Homo sapi
18	319	14.1	5322	6	AX405756	Sequence
19	273.4	12.1	61103	2	AC131276	Homo sapi
20	191.2	8.4	193437	2	AC126667	Rattus no
21	144	6.3	52684	2	AC014796	Drosophil
22	144	6.3	174712	3	AC010041	Drosophil
23	144	6.3	195148	3	AC010019	Drosophil
24	144	6.3	282267	3	AE003555	Drosophil
25	129.2	5.7	2174	10	BC021530	Mus muscu
26	128.4	5.7	2015	9	AB051127	Macaca fa
27	127.6	5.6	4573	6	AX409002	Sequence
28	127.6	5.6	4573	9	HUMOREKGLP	Human mRNA
29	126	5.6	4610	6	AX078294	Sequence
30	126	5.6	4698	6	AX281784	Sequence
31	123.4	5.4	1383	6	AX118897	Sequence
32	123.4	5.4	1803	9	AF193052	Homo sapi
33	123.4	5.4	2604	9	AK027652	Homo sapi
34	123.4	5.4	3169	9	BC012125	Homo sapi
35	123.4	5.4	3246	9	AB040120	Homo sapi
36	122.6	5.4	1895	9	BC027884	Homo sapi
37	119.2	5.3	1755	9	BC015770	Homo sapi
38	119.2	5.3	55577	2	AC014109	Homo sapi
39	119.2	5.3	170071	3	AC093497	Drosophil
40	119.2	5.3	194979	3	AC091209	Drosophil
41	119.2	5.3	295377	3	AE003532	Drosophil
42	118	5.2	150	9	HUMDSF04M5	Human HepG2
43	116.2	5.1	1931	10	BC028990	Mus muscu
44	115.8	5.1	3060	10	BC006731	Mus muscu
45	110.6	4.9	1888	6	AX061633	Sequence

ALIGNMENTS

RESULT 1
AX465588
LOCUS AX465588
DEFINITION Sequence 1 from Patent WO0216939.
ACCESSION AX465588
VERSION AX465588.1
KEYWORDS GI:21899891
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mack D., Gish, K.C. and Wilson, K.E.
TITLE Methods of diagnosis of cancer and screening for cancer modulators
JOURNAL Patent: WO 0216939-A 1 28-FEB-2002;

AX465588 3461 bp DNA linear PAT 16-JUL-2002

FEATURES	EUS Biotechnology, Inc. (US)
source	Location/Qualifiers
	1..3461
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/note="prustate/breast/ovarian/bladder cancer protein
CDS	human BCR4 cDNA
	138...2405
	/note="human BCR4"
	/codon_start=1
	/protein_id="CAD42374.1"
	/db_xref="GI:21899892"
	/translation="MARKLSVILLITPALSVTNPILHEIKAAAFPTQTEKISPWNESGIVNDLAISTROYHQLOFVRYGNNSSUVEGRFKLLONGIGLIKRIIHHHDHSHDHEDSHEDSHDHHHSNHAASQGNKRKAICLPDSDSSGCKDPKN SQKGAPRHPESHARRNRKDVSASEVTSTVYNISEGTHTLETETPRPKLPKDVSSTSPSVTSKSRSLAGRKNTSYSEPRKGPMTYSRNTNENPQCFAASKLLTSGM GIOVPLNATLEPCALINOIDARSCLLHTSEKKAEIPPKTYSLQIAWVGGLAISITFSLSLGLVFLPYLMNRVPFKFLLSELVALAVGTLSGDATFLHLPHSHASHHHSHSE EPMEMKRGFLPSHLSNQNI EESAYFDSTWKGLTAGLGLEYENLFEBHVTLTKQFKDK KKNKKPPENDDOVEIKKLQSYEQSLNEEKVDTDRTKEGVLADRSORPSPHFSDQQ PAVLEEVEYIAHANPEYNYEYVRCKNKNSHPHDTIGQSDOLI HHHDHVFHTLUH HHHNHHPHSHLPESYSRELRKDGAVATLAWMV IMGDLGHNFSGINGLA IGAATPEGLSS GISTSVAFCHSHURHELJGDGFVALLKAGMTYKOAVLYNALSAMLAYLGMATGLPIGHYA ENYSMMIFALTACLFWYVALVDMPVKMLINDASDHGCSRRCYFFLDNAGMELLEGFGIML LISIEBKIVFRNF"
BASE COUNT	1042 a 686 c 728 g 1005 t
ORIGIN	
	Query MultCh 100.0% Score 2268; DB 6; Length 3461;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
UY	1 ATGCGAGAGTAATCTGTAATCTTGATCTCGACCTTTGCCCTCTCTCTGTACAAAATCC 60
D	148 ATGCGAGGAAGTTATCTGTAATCTTCATCCTGACCTTTGCCCTCTCTCTGTCAATAATCCC 197
UY	61 CTTTATCAACTAAAGCAGCTGGCTTTCCCCAGACCACCTCAGAAAAATTACTTCCAATTCG 120
D	198 CTTTCATCAACTAAAGCAGCTGGCTTTCCCCAGACCACCTCAGAAAAATTACTTCCAATTCG 257
UY	121 GAATCTGCCATTAATGTTGACTTTGGCAATTTCCACACGGCAATATCATCTACCAACAGCTT 180
D	258 GAATCTGGCATTAATGTTGACTTTGGCAATTTCCACACGGCAATATCATCTACCAACAGCTT 317
UY	181 TTCTACCGCTATGAGAAAAATAATCTTTCTGCTAGTTGAAGGGTTCAGAAAAATTACTTCAA 240
D	318 TTCTACCGCTATGAGAAAAATAATCTTTCTGCTAGTTGAAGGGTTCAGAAAAATTACTTCAA 377
UY	241 AATATAGGCATAGATAAGATTAAAGAATCCATATACACCATGACCACGACCATCATCA 300
D	378 AATATAGGCATAGATAAGATTAAAGAATCCATATACACCATGACCACGACCATCATCA 437
UY	301 GAGCATGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCCAC 360
D	438 GACCACGACCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCCAC 497
UY	361 GAGCATCACTCTGACCATGATCATCACTCTCACCATTAATCATGCTGCTTCTGTATAAAT 420
D	498 GAGCATCACTCTGACCATGATCATCACTCTCACCATTAATCATGCTGCTTCTGTATAAAT 557
UY	421 AAGCGAAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTAGAAC 480
D	558 AAGCGAAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTAGAAC 617
UY	481 AGCCAGGGGAAGGAGCTCACTGACCCAGAACATGCCAGTGGTAGAAGGAATCTCAAGGAC 540
D	618 AGCCAGGGGAAGGAGCTCACTGACCCAGAACATGCCAGTGGTAGAAGGAATCTCAAGGAC 677
UY	541 AGTGTAGTGTAGTGAATGACCTCACTGCTACAACACTGTCTCTGTAAGAGAAATCTCAAC 600
D	678 AGTGTAGTGTAGTGAATGACCTCACTGCTACAACACTGTCTCTGTAAGAGAAATCTCAAC 717

[illegible]

Db	318	TTCTACCGTATYGGAGAAATAATCTTTGTGAGTGTGAGGGTTACAAAAATAC	TCAA	377
QY	241	AATATAGGCATAGATYAAGATTAAAGAATCCATATACAGCATGACCAACGACCATCACTCA	300	
Db	378	AATATAGGCATAGATYAAGATTAAAGAATCCATATACAGCATGACCAACGACCATCACTCA	437	
QY	301	GATCACGAGCATCACTCAGACCATGAGCGTCACATCAGCATGAGCATCACTCAGACATC	360	
Db	418	GATCACGAGCATCACTCAGACCATGAGCGTCACATCAGCATGAGCATCACTCAGACATC	497	
QY	361	GAGCATYACCTCTGACCATGATCATCACTCTCAGCATTAATCATGCTGCTTCTGCTAAAAT	420	
Db	498	GAGCATYACCTCTG-----ACCATTAATCATGCTGCTTCTGCTAAAAT	539	
QY	421	AGCGAAAGCTCTTTTCGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTAAAC	480	
Db	540	AGCGAAAGCTCTTTTCGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTAAAC	599	
QY	481	AGCTAGGGGAAGGAGCTCACCGACCAGAACATGCCAGTGGTGAAGAAGAAATGTCAGGAC	540	
Db	600	AGCCAGCGGAAGCAGCTCACCGACCAGAACATGCCAGTGGTGAAGAAGAAATGTCAGGAC	659	
QY	541	AGTGTTAGTGTAGTGAAGTGAAGTCAACTGTGTACAAGACTGTCTTGAAGAAATCTCAC	600	
Db	660	AGTGTAGTGTAGTGAAGTGAAGTCAACTGTGTACAAGACTGTCTTGAAGAAATCTCAC	719	
QY	601	TTTTCTAGAGCAATFAGAGACTTCAAGACCTGGAAAACTCTTCCCAAGAGATCTAAGCAGC	660	
Db	720	TTTTCTAGAGCAATFAGAGACTTCAAGACCTGGAAAACTCTTCCCAAGAGATCTAAGCAGC	779	
QY	661	TCCACTCCACCAGTGTACATCAAAAGCCGGGTGAGCGGCTGGCTGTGTAGSAAAAACA	720	
Db	780	TCCACTCCACCAGTGTACATCAAAAGCCGGGTGAGCGGCTGGCTGTGTAGSAAAAACA	839	
QY	721	AATGAATCTGTAGTGAGCCCCCAAAAGGCTTTATGTATTTCCAGAAACACAATGAAAT	780	
Db	840	AATGAATCTGTAGTGAGCCCCCAAAAGGCTTTATGTATTTCCAGAAACACAATGAAAT	899	
QY	781	CCTCAGGAGTGTTCATATGATCAAAAGCTACTGACATCTCATGGATGGCATCCAGGTT	840	
Db	900	CCTCAGGAGTGTTCATATGATCAAAAGCTACTGACATCTCATGGATGGCATCCAGGTT	959	
QY	841	CCGCTGAATGCACAGAGTTCAACTATCTGTGTCCAGGCATCATCAACCAAAATTCATGCT	900	
Db	960	CCGCTGAATGCACAGAGTTCAACTATCTGTGTCCAGGCATCATCAACCAAAATTCATGCT	1019	
QY	901	AGATCTTTGTCTGATTCATACAAGTGAAGAAGGCTGAAATCCATCCAAAGACCTATTCA	960	
Db	1020	AGATCTTTGTCTGATTCATACAAGTGAAGAAGGCTGAAATCCATCCAAAGACCTATTCA	1079	
QY	961	TTACAAATAGCCGTGGTGGTGGTTTATAGCCATTTCCATCATCAGTTTCTGCTGCTG	1020	
Db	1080	TTACAAATAGCCGTGGTGGTGGTTTATAGCCATTTCCATCATCAGTTTCTGCTGCTG	1139	
QY	1021	CTGGGGCTTATCTTTAGTGCCCTCFCATGAATCGGGTCTTTTTCAAATTTCTCTCAGTTTC	1080	
Db	1140	CTGGGGCTTATCTTTAGTGCCCTCFCATGAATCGGGTCTTTTTCAAATTTCTCTCAGTTTC	1199	
QY	1081	TTTGTGSCATGSCCGTGGGACTTTTGTAGTGGTGATGCTTTTTTACAGCTTCTTCCACAT	1140	
Db	1200	TTTGTGSCATGSCCGTGGGACTTTTGTAGTGGTGATGCTTTTTTACAGCTTCTTCCACAT	1259	
QY	1141	TCTCATGCAAGTCACCAACATAGTCATAGCCATGAAGAACAGCAATGGAAATGAAAGA	1200	
Db	1260	TCTCATGCAAGTCACCAACATAGTCATAGCCATGAAGAACAGCAATGGAAATGAAAGA	1319	
QY	1201	GGACCAATTTTTCAGTCACTGTCTTCTCAAAACATAGAAAGAGTGCCTATTTTGTATTC	1260	
Db	1320	GGACCAATTTTTCAGTCACTGTCTTCTCAAAACATAGAAAGAGTGCCTATTTTGTATTC	1379	
QY	1261	ACGTGGAGGGTCTACAGCTCTAGAGGCGCTGTATTTTCATGTTTCTGTGTGAACATGTC	1320	

Db	1380	ACGTGGAAAGGCGCTAAACAGCGCTCTAGGAGGCGCTGATTATTCATGTTCTTGTTCGAACATGTC	1439
Qy	1321	CTCACATTGATCAAAACAATTTAAAGATTAAGAACAAAAACAATCAGAAGAAACCTCAAAAT	1380
Db	1440	CTCACATTGATCAAAACAATTTAAAGATTAAGAGAGAAAACAATCAGAAGAAACCTGAAAAT	1499
Qy	1381	GATGATGATGGGAGATTTAAGAAGCAGTTGTGCCAGTATGAATCTCAACTTTTCAACAAT	1440
Db	1500	GATGATGATGGGAGATTTAAGAAGCAGTTGTGCCAGTATGAATCTCAACTTTTCAACAAT	1559
Qy	1441	GAGGAGAAAGTAGATACAGATGATCCNACTGAAGGCTATTTACGAGCAGACTCAACAAG	1500
Db	1560	GAGGAGAAAGTAGATACAGATGATCCNACTGAAGGCTATTTACGAGCAGACTCAACAAG	1619
Qy	1501	CCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTCTTGGAAAGAAAGAGGTCATGATAGCT	1560
Db	1620	CCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTCTTGGAAAGAAAGAGGTCATGATAGCT	1679
Qy	1561	CATGCTCATCCACAGGAGTCTACAATGAATATGTACCCAGAGGTCGAAGAATAAATGC	1620
Db	1680	CATGCTCATCCACAGGAGTCTACAATGAATATGTACCCAGAGGTCGAAGAATAAATGC	1739
Qy	1621	CATTACATTTCCACGATACACITCGCCAGTCAGACGATCTCATTTACCCACCATCATCAC	1680
Db	1740	CATTACATTTCCACGATACACTCGCCAGTCAGACGATCTCATTTACCCACCATCATGAC	1799
Qy	1681	TACCATCATATTCCTCCATCATCACCAACCACCAACCATCTCCACAGTCACAGCCAG	1740
Db	1800	TACCATCATATTCCTCCATCATCACCAACCACCAACCATCTCCACAGTCACAGCCAG	1859
Qy	1741	CGCTACTCTCGGAGGAGCTGAAGATGCGGGCTGCCACTTTGGCCCTGGATGGTGATA	1800
Db	1860	CGCTACTCTCGGAGGAGCTGAAGATGCGGGCTGCCACTTTGGCCCTGGATGGTGATA	1919
Qy	1801	ATGGGTGATGGCCTGGACAATTTACCGGATGGCCTAGCAATTTGGTGCTCTTTACTGAA	1860
Db	1920	ATGGGTGATGGCCTGGACAATTTACCGGATGGCCTAGCAATTTGGTGCTCTTTACTGAA	1979
Qy	1861	GGCTTATCAAGTGGTTTAAGTACTCTGTGGTGTGTCTGTCTATGAGTTGCCCTCATGAA	1920
Db	1980	GGCTTATCAAGTGGTTTAAGTACTCTGTGGTGTGTCTGTCTATGAGTTGCCCTCATGAA	2039
Qy	1921	TTAGGTGACTTTGCTGTCTTCTACTAAAGCGTGGCATGACCGTTTAAGCAGGCTGCTCTTAT	1980
Db	2040	TTAGGTGACTTTGCTGTCTTCTACTAAAGCGTGGCATGACCGTTTAAGCAGGCTGCTCTTAT	2099
Qy	1981	AATGCATTTGACCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTCAT	2040
Db	2100	AATGCATTTGACCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTCAT	2159
Qy	2041	TATGCTCAAAATGTTTCTATGTGGATATTTGACATTTACTGCTGGCTTATTCATGATGCTTT	2100
Db	2160	TATGCTCAAAATGTTTCTATGTGGATATTTGACATTTACTGCTGGCTTATTCATGATGCTTT	2219
Qy	2101	GCTCTGGTTGATATGCTACTGAAATGCTGCAACATGATGCTAGTGACCATGGATGTAGC	2160
Db	2220	GCTCTGGTTGATATGCTACTGAAATGCTGCAACATGATGCTAGTGACCATGGATGTAGC	2279
Qy	2161	CGCTGGGGTATTTCTTTTACAGAAATGCTGGATGCTTTTGGGTTTGGAAATTTATGTTA	2220
Db	2280	CGCTGGGGTATTTCTTTTACAGAAATGCTGGATGCTTTTGGGTTTGGAAATTTATGTTA	2339
Qy	2221	CTTATTTCCATATTTCAACATAAATCTGTTTCTGATTAATTTCTAG	2268
Db	2340	CTTATTTCCATATTTCAACATAAATCTGTTTCTGATTAATTTCTAG	2387

RESULT 4			
AX207205			
LOCUS	AX207205	3461 bp	DNA
DEFINITION	Sequence 1 from Patent WO0153178.		
ACCESSION	AX207205		
VERSION	AX207205.1	GI:15394960	
			PAT 30-AUG-2001

Db	897	CTCCAAAGACCTATTTCTTTACAAATAGCCTGGCTTGGTGGCTATPAGCCATTTCTCCA TCA	956
QY	1004	TCAGTTTCTCTCTCTCTCTGGGGGTATATTTTACTGTGCTCTCATGAATCGGTTGTTTTC	1063
Db	957	TCAGTTTCTCTCTCTCTCTGGGAGCTCATCTTGTGTGCCACTCATGAACCGGTATATTTTTC	1016
QY	1064	AAATTTCTCTCAGTTTCTCTTCTGTCACATTCGCGCTTGGAGCTTTGAGTGTGTATCTCT	1123
Db	1017	AGTTTCTCTCTGCTGAGCTTCTCTCTGCGGTGGCGCTCGGAACGCTGAGTGGCGATCTCTCT	1076
QY	1124	TACACCTTTCTTCCACATTTCTCATCATCAAGTCCACCACCATAGTCATAGCCATCAAGAACTCAG	1183
Db	1077	TACATCTTTCTCCACACTCTCATCTCAAGTCAATCAGCACACTCATAGCCATGAAGAGCCAG	1136
QY	1184	CAATGGAATGAAGAAGGACACACTTTTTCAGTCAATCTGTCTTCTCAAAACATAGAAAGAA	1243
Db	1137	CGATGGAATGAAGAAGGCCCTCTGTTCAGCCACCTGTCTGGCTCAGATAATAGAAAGAA	1196
QY	1244	GTGGCTATTTTGTATTCACAGTGGAAAGGCTCTAACAGTCTTAGAGGCCCTGTATTTTCATGT	1303
Db	1197	GCTCTCTATTTGTATTCACAGTGGAAAGGCTCTGACGGCTCTAGGGGCTTATATTTCTATGT	1256
QY	1304	TTCTTGTGTGAACATGTCTCTCACATTTGATCAAAACAATTTAAAGATGAAGAGAAAGAAATC	1363
Db	1257	TTCTTGTGTGAACACGTACTCACACTGATCAAGCAATTTTAAAGATAAGAAAGAAAGAAATC	1316
QY	1364	AGAAAGAACCTGAAATGATGATGTCTGAGATTAAGANGCAGTTGTTCCAAATTTATCAAT	1423
Db	1317	AAAGAAACCTGAAATGATGAGGATGTGAGAGGACGAAGCAGCGTGTCCAAATACACACT	1376
QY	1424	CTCAACTTTTCACAAATGAGGAGAAAGTATAGATACAGATGATCGAACTGAAGGCTATTTTAC	1483
Db	1377	CTCAACTTTTCTCTCAAAATGAAGAGAGGTGGACCCAGGGGAACGACCTGGAAGACTATCTGC	1436
QY	1484	CAGCAGACTCACAAAGAGCCCTCCACTTTGATTTCTCAGCGCTCCACTCTTTGAAGAAG	1543
Db	1437	GAGCCGACTCCCAAGAGCCCTCCCTTTGTATTCGCCAGCGCCAGCATGTGTGAAGAGG	1496
QY	1544	AAGAGCTATGATAGTCTATGCTCATPCCACAGGAAGTCTACAATGAATATGTACGCCAGAG	1603
Db	1497	AAGAGTTCATGATAGGCCATAGGCCATGACACCCACAAGAAGTCTTCAATGAATATCTGCCACGG	1556
QY	1604	GGTCAAGAATAAATGCCATTTCCAGATTCACACTCGGCCAGTTCAGATGATCTCA	1663
Db	1557	GCTCCAGAACAATGCCATTTCCAGATTCACACTTCGAGATGAGCTGGGCCAGTCCGAGGAGCTCA	1616
QY	1664	TTCAACCACCATATGACTACCATCATATTTCTCCATCATCACCCACCAACCAACCATC	1723
Db	1617	TCCACACCATATCAGACTACCATCATATTTCTGACCATTCCTGACCCACCCACCAAGCACCC	1676
QY	1724	CTCACATCTCAGCCAGCCCTACTCTCGGGAGGAGCTGAAAGATGCGGGGTGCGGCACAT	1783
Db	1677	CTCACAGCCACAGCCAGCCCTACTCTCAGAGAGAGCTGAAGAGCGCCGGCATTTGCCACAT	1736
QY	1784	TGCGCTGGATGTGATTAATGGGTGATGCGCTGCACAAATTTTCAGGATGCGCTTAGTAATTC	1843
Db	1737	TGCGCTGGATGTGATCATGCGCGAGGAGCTGCAACAATTTTCAGTGAGCGGCTTTGCTATTG	1796
QY	1844	GTGCTGCTTTTACTTGAAGGCTTATCAAGTGGTTTAAAGTACTTCTGTTGCTGTCTCTGTCTG	1903
Db	1797	GTGCTGCTTTTCAACCGAGGTTTGTCCAGTGGGCTTAAGCACCTCTGTGCGCTGTCTCTGTCTG	1856
QY	1904	ATCAGTTTGCCTCATCAATTTAGTGACTTTTGTCTGTCTTACTTAAAGGCTGCGATGACCGTTA	1963
Db	1857	ATCAACTGCCTCATGAACCTAGTGACTTTTGTCTGTCTTTGCTTAAAGGCTGGCATGACGTCA	1916
QY	1964	AGCAGCTCTCTCTTATAATGCATTTGTCTCAGCCATGCTGGCGTATCTTTGAAATGGTAACAG	2023
Db	1917	AUCAGGCTCTGCTACTAATGCTCTGTCTCAGCCATGTTGCGCTACCTTTGGAAATGCAATAG	1976
QY	2024	GAATTTTCATTCGTCTAATTCGTGAAATGTTTCTCTATGTGCGATATTTTGGCACTTATCTCTC	2083

Db	1977	GGATATTCAATCGGGCAATTATGCAAAAAATATTTCTATGTGGATATTTCCGACTTCACATGCCG	2036
QY	2084	GCATTATTCAATGATGTGCTCGTTTGATATGGTACCTGAAATGCCTGCACAATGATGCCTA	2143
Db	2037	GCCTTTTCATGATGTCGCTCTGCTTGACATGGTGCCTGAGATGTTGCACAATGATGCCTA	2096
QY	2144	GTGACCATGGATGATGCCCGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGG	2203
Db	2097	GTGATCAGCGATGCAGCCGCTGGGGAATATTTCTTCCTCGAGAATGCTGGGATACTCTCG	2156
QY	2204	GTTTTGGAATTATGTTACTATTATTTCCATATTTGAACATAAAATCGTGTTTCCGTATAAAAT	2263
Db	2157	GTTTTGGAATTATGTTACTCAATTTCCATATTTGAGCATAAAAATTTGTTTTCGTATAAAAT	2216
QY	2264	TCTA 2267	
Db	2217	TCTA 2220	
RESULT 6			
LOCUS	176891	176891	3310 bp DNA linear PAT 03-APR-1998
DEFINITION		Sequence 1 from patent US 5693465.	
ACCESSION		176891	
VERSION		176891.1	GI:3013045
KEYWORDS			
SOURCE		Unknown.	
ORGANISM		Unknown.	
REFERENCE		Unclassified.	
AUTHORS		1 (bases 1 to 1310)	
TITLE		Manning,D.Lockwood., Nicholson,R.Ian., Gee,J.Margaret.Wendy. and	
JOURNAL		Green,C.Douglas.	
FEATURES		Methods for predicting the behaviour of breast tumours	
		Patent: US 5693465-A 1 02-DEC-1997;	
		Location/Qualifiers	
	source	1..1310	
		/organism="unknown"	
BASE COUNT	354 a	271 c	287 g 398 t
ORIGIN			
Query Match	55.7%;	Score 1263;	DB 6;
Best Local Similarity	99.8%;	Pred. No. 1.8e-302;	Length 1310;
Matches 1285;	Conservative 0;	Mismatches 0;	Indels 2; Gaps 2;
QY	982	GTTTTTATAGCCATTTCCATCATCAGTTTCCTGTCTCTGCTGGGGGTTATCTTAGTGCCT	1041
Db	1	GTTTTTATAGCCATTTCCATCATCAGTTTCCTGTCTCTGCTGGGGGTTATCTTAGTGCCT	60
QY	1042	CTCATGAATCGGGTGTTTTTCAAATTTCTCTGAGTTTCTTTGGGCACTGCCGTTGGG	1101
Db	61	CTCATGAATCGGGTGTTTTTCAAATTTCTCTGAGTTTCTTTGGGCACTGCCGTTGGG	120
QY	1102	ACTTTGAGTGGTGATGCTTTTTTACACCTTCTCCACATTTCTCATGCAAGTCACCCAT	1161
Db	121	ACTTTGAGTGGTGATGCTTTTTTACACCTTCTCCACATTTCTCATGCAAGTCACCCAT	180
QY	1162	AGTCATAGCCATGAAGAACCCAGCAATGGAATGAAAGAGGACCACCTTTTCAGTCATCTG	1221
Db	181	AGTCATAGCCATGAAGAACCCAGCAATGGAATGAAAGAGGACCACCTTTTCAGTCATCTG	240
QY	1222	TCTTCTCAAACATAGAGAAGTGCCCTATTTTGATTTCCACGTGGGAAGGCTCTAACAGCT	1281
Db	241	TCTTCTCAAACATAGAGAAGTGCCCTATTTTGATTTCCACGTGGGAAGGCTCTAACAGCT	300
QY	1282	CTAGAGGCCCTGATTTTCATGCTTTCTTTGTTGAACATGCTCTCACATTTGATCAACAATTT	1341
Db	301	CTAGAGGCCCTGATTTTCATGCTTTCTTTGTTGAACATGCTCTCACATTTGATCAACAATTT	360
QY	1342	AAAGATAAGAAGAAAAAGAAATCAGAAGAAACCTCGAAATGATGATGATGTGGAGATTAAAG	1401
Db	361	AAAGATAAGAAGAAAAAGAAATCAGAAGAAACCTCGAAATGATGATGATGTGGAGATTAAAG	420
QY	1402	AAGCAGTTGTTCCAAGTATGAATCTCAACTTTTCAACAAATGAGGAGAAAGTAGATACAGAT	1461

db	421	AAGCAGTTCTCCAAAGTATCAATCTCAACTTTCAACAATGAGCAAGAAAGTACATATACAGAT	480
QY	1462	GATCGAATCTGAAGGGTATTTATGAGCAGACTCAAGAGGCTCTCCACTTTTGATTTCTGAG	1521
db	481	GATCGAATCTGAAGGGTATTTATGAGCAGACTCAAGAGGCTCTCCACTTTTGATTTCTGAG	540
QY	1522	CAGCCTCTGAGTCTTCSAAGAAGAACAGGTCTCATGATAGCTCATGCTTCATCTCACAGTAAGTTC	1581
db	541	CAGCCTCTGAGTCTTCSAAGAAGAACAGGTCTCATGATAGCTCATGCTTCATCTCACAGTAAGTTC	600
QY	1582	TACAAATGAATATGTACCCAGAGGGTGCAGAAATAAATGCCATTACATTTCCACCAATACA	1641
db	601	TACAAATGAATATGTACCCAGAGGGTGCAGAAATAAATGCCATTACATTTCCACCAATACA	660
QY	1642	CTCGGCTGAGTCTAGAGCATCTCATTCACCAACCATCATGACTACCATCATATTTCTTCATCAT	1701
db	661	CTCGGCTGAGTCTAGAGCATCTCATTCACCAACCATCATGACTACCATCATATTTCTTCATCAT	720
QY	1702	CACCACCAACCAACCAACCTCTCACAGTTCACAGCCAGCGCTACTCTCGGAGCAGCGTCTG	1761
db	721	CACCACCAACCAACCAACCTCTCACAGTTCACAGCCAGCGCTACTCTCGGAGCAGCGTCTG	780
QY	1762	AAAGATCGCGGCTGCCACATTTGCGCTCGATGTGATAATGGGTGATGGCTTCGACAAAT	1821
db	781	AAAGATCGCGGCTGCCACATTTGCGCTCGATGTGATAATGGGTGATGGCTTCGACAAAT	840
QY	1822	TTACGGGATGGCTTACCAATTTGGTGCTCTCTTTACTGAAGGCTTATCAAGTGTGTTAAGT	1881
db	841	TTACGGGATGGCTTACCAATTTGGTGCTCTCTTTACTGAAGGCTTATCAAGTGTGTTAAGT	900
QY	1882	ACTTCTGTTGCTGTCTGTCTCATGAGTTGGCTCATGAATTAGGTGACTTTTGTCTTTCTTA	1941
db	901	ACTTCTGTTGCTGTCTGTCTCATGAGTTGGCTCATGAATTAGGTGACTTTTGTCTTTCTTA	960
QY	1942	CTAAAGGCTGGCATGACCGTTTAAAGCAGGCTGTCTTTTATATGCAATTTCTCAATGCTG	2001
db	961	CTAAAGGCTGGCATGACCGTTTAAAGCAGGCTGTCTTTTATATGCAATTTCTCAATGCTG	1020
QY	2002	GGTATCTTTGCAATGGCAACAGGAATTTTCATTTGGTGTATATGCTGAAATGTTTCTATG	2061
db	1021	GGTATCTTTGCAATGGCAACAGGAATTTTCATTTGGTGTATATGCTGAAATGTTTCTATG	1080
QY	2062	TGGATATTTGCAATTTGCTGTGCTTATTCATGATGTTGCTGTCTGTTCATATGCTACCT	2121
db	1081	TGGATATTTGCAATTTGCTGTGCTTATTCATGATGTTGCTGTCTGTTCATATGCTACCT	1140
QY	2122	CAATGCTGCAATGATGCTAGTACCATGATCTAGCCGCTGGGCTATTTCTTTTCTTA	2181
db	1141	CAATGCTGCAATGATGCTAGTACCATGATCTAGCCGCTGGGCTATTTCTTTTCTTA	1200
QY	2182	CAGAACTGGGATGCTTTTGGGTTTTGGAAATATGTTACATTTTTCATATTTCAAT	2241
db	1201	CAGAACTGGGATGCTTTTGGGTTTTGGAAATATGTTACATTTTTCATATTTCAAT	1259
QY	2242	AAATCTGTTTCTGATATAAATTTCTAG	2268
db	1260	AAATCTGTTTCTGATATAAATTTCTAG	1285
RESULT 7	176892		
LOCUS	176892	2404 bp	DNA linear PAT 03-APR-1998
DEFINITION	Sequence 3 from patent US 5693465.		
ACCESSION	176892		
VERSION	176892.1	GI:3013046	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2404)		
AUTHORS	Manning, B. Lockwood., Nicholson, R. Ian., Gee, J. Margaret, Wendy. and Green, C. Douglas.		

1898 TCTGTGTCATGAGTTGGCTCATGCAATTAAGCTGTTCTTACTAAAGGCTGACATGA 1957
|||||
781 TCTGTGTCATGAGTTGGCTCATGCAATTAAGCTGTTCTTACTAAAGGCTGACATGA 840
|||||
1958 CCGTTTAAGCAGCGTCCTTTTAAATGATGATGTCAGCCATGCTGGCGTATCTTGGAAATGG 2017
|||||
841 CCGTTTAAGCAGCGTCCTTTTAAATGATGATGTCAGCCATGCTGGCGTATCTTGGAAATGG 900
|||||
2018 CAACAGGAAATTTTCATGTCATGTCATGTCGAAATGTTTCTATGTGATATTTTCACATTA 2077
|||||
901 CAACAGGAAATTTTCATGTCATGTCATGTCGAAATGTTTCTATGTGATATTTTCACATTA 960
|||||
2078 CTGCTGGCTTATTCATGTCATGTCGTCGTTGATATGCTACCTGAAATGCTGCACAAATG 2137
|||||
961 CTGCTGGCTTATTCATGTCATGTCGTCGTTGATATGCTACCTGAAATGCTGCACAAATG 1020
|||||
2178 ATGCTAGTCAGCATGATGTCAGCCGCTGGGGGTATTTCTTTTACAGAATGCTGGGATGC 2197
|||||
1021 ATGCTAGTCAGCATGATGTCAGCCGCTGGGGGTATTTCTTTTACAGAATGCTGGGATGC 1080
|||||
2198 TTTTGGCTTTTGGAAATGATGTTACTTATTTCCATATTTGAACATAAATCGTTTTCCTA 2257
|||||
1081 TTTTGGCTTTTGGAAATGATGTTACTTATTTCCATATTTGAACATAAATCGTTTTCCTA 1140
|||||
2258 TAAATTTCTAG 2268
|||||
1141 TAAATTTCTAG 1151
|||||

AP001158 160170 bp DNA linear HTG 13-JUL-2000
Homo sapiens chromosome 18 clone RP11-701C9 map 18q12, WORKING
DRAFT SEQUENCE, 15 unordered pieces.
AP001158
AP001158.3 GI:9188478
HTG: HTGS_PHASET1; HTGS_DRAFT.
Homo sapiens DNA, clone: RP11-701C9.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 160170)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seon,P.,
Fujiyama,A., Yada,T., Toki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 160,170 genomic DNA of 18q12
Published Only in DataBase (2000)
2 (bases 1 to 160170)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seon,P.,
Fujiyama,A., Yada,T., Toki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (10-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8118726.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: Humbrat118
Center clone name: RP11-701C9
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.940329
Consensus quality: 155820 bases at least Q40
Consensus quality: 157459 bases at least Q30
Consensus quality: 158091 bases at least Q20
Insert size: 158770; sum-of-contigs

Quality coverage: 10.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
15 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1
29590 contig of 29590 bp in length
29691 contig of 20216 bp in length
49906 contig of 19114 bp in length
69120 contig of 14972 bp in length
84192 contig of 15598 bp in length
99890 contig of 10477 bp in length
110467 contig of 7924 bp in length
118491 contig of 7858 bp in length
126449 contig of 8409 bp in length
134958 contig of 5995 bp in length
141053 contig of 4767 bp in length
149439 contig of 3419 bp in length
154173 contig of 4634 bp in length
157616 contig of 3343 bp in length
160170 contig of 2454 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
29590: contig of 29590 bp in length
29691: gap of 100 bp
49906: contig of 20216 bp in length
69120: gap of 100 bp
84192: contig of 19114 bp in length
99890: gap of 100 bp
110467: contig of 14972 bp in length
118491: gap of 100 bp
126449: contig of 7858 bp in length
134958: gap of 100 bp
141053: contig of 8409 bp in length
149439: gap of 100 bp
154173: contig of 5995 bp in length
157616: gap of 100 bp
160170: contig of 4634 bp in length
160170: contig of 3343 bp in length
160170: gap of 100 bp
160170: contig of 2454 bp in length.
Location/Qualifiers
1. 160170
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-701C9"
1. 29590
/note="assembly_fragment"
29691. .49906
50007. .69120
misc_feature
misc_feature
misc_feature

FEATURES
Source

```
/note="assembly_fragment"
69221..84192
/note="assembly_fragment"
84293..99890
/note="assembly_fragment"
99991..110467
/note="assembly_fragment"
110568..118491
/note="assembly_fragment"
118592..126449
/note="assembly_fragment"
126550..134958
/note="assembly_fragment"
135059..141053
/note="assembly_fragment"
141154..145920
/note="assembly_fragment"
146021..149439
/note="assembly_fragment"
149540..154173
/note="assembly_fragment"
154274..157616
/note="assembly_fragment"
157717..160170
/note="assembly_fragment"
BASE COUNT 46219 a 33232 c 33436 g 45883 t 1400 others
ORIGIN
Query Match 34.8%; Score 789; DB 2; Length 160170;
Rest local Similarity 100.0%; Pred. No. 1.4e-184;
Matches 789; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGAGGAGTTATCTGTAATCTTGATCTTGACCTTTGCGCTCTCTCTCAACAAATCC 60
Db 3182 ATGCGAGGAGTTATCTGTAATCTTGATCTTGACCTTTGCGCTCTCTCTCAACAAATCC 31882
QY 61 CTTTCATCACTAAAGCAGCTGCTTTCCGCCGACCACTGAGAAATTAAGTCGAATTCG 120
Db 3188 CTTTCATCACTAAAGCAGCTGCTTTCCGCCGACCACTGAGAAATTAAGTCGAATTCG 31942
QY 121 GAATCTGGCAATTAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACAAAGCTT 180
Db 31943 GAATCTGGCAATTAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACAAAGCTT 32002
QY 181 TTTTACCGCTATCGAGAAATAATATCTTTGTCTAGTTTGAAGGTTTCAGAAATTAATCTCAA 240
Db 32003 TTTTACCGCTATCGAGAAATAATATCTTTGTCTAGTTTGAAGGTTTCAGAAATTAATCTCAA 32062
QY 241 AATATAGGCATACATGAATTAAGAAATCCATATACACCATGACCAACCATCATCTCA 300
Db 32063 AATATAGGCATACATGAATTAAGAAATCCATATACACCATGACCAACCATCATCTCA 32122
QY 301 GACCATGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAT 360
Db 32123 GACCATGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAT 32182
QY 361 GAGCATCACTCTGACCATGATCATCACTCTCAACATAATCATGCTCTCTGTAATAAT 420
Db 32183 GAGCATCACTCTGACCATGATCATCACTCTCAACATAATCATGCTCTCTGTAATAAT 32242
QY 421 AAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGTTAAGATCTTGAAC 480
Db 32243 AAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGTTAAGATCTTGAAC 32302
QY 481 ACCGAGGGAAGAGCTCTCCCGACACAGAAATGCGAGTGGTGAAGGAATCTCAAGAC 540
Db 32303 ACCGAGGGAAGAGCTCTCCCGACACAGAAATGCGAGTGGTGAAGGAATCTCAAGAC 32362
QY 541 ACTGTTAGTCTAGTGAAGTACCTCAACTGTGTACACACTCTCTCTCAAGCAATTCAC 600
Db 32363 ACTGTTAGTCTAGTGAAGTACCTCAACTGTGTACACACTCTCTCTCAAGCAATTCAC 32422
QY 601 TTTTCTAGACACAATAGAGACTCCAGACCTCGAAAACCTTTTCCCAAAATCTTAACACAC 660
```

```
|||||
Db 32423 TTTCTAGACACAATAGAGACTCCAGACCTGGAACACTTTCCCAAAAGATCTTAAGCAC 32482
QY 661 TCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGCTGGCTGTAGGAAACA 720
Db 32483 TCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGCTGGCTGTAGGAAACA 32542
QY 721 AATGAATCTGTGAGTCAGCCCGAAAGGCTTTATGTATTTCAGAAACAAATGAAAT 780
Db 32543 AATGAATCTGTGAGTCAGCCCGAAAGGCTTTATGTATTTCAGAAACAAATGAAAT 32602
QY 781 CCTCAGGAG 789
Db 32603 CCTCAGGAG 32611
|||||
RESULT 10
AC091060/c
AC091060/c
AC091060
AC091060.10 GI:16646826
HTG.
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
1 (bases 1 to 221941)
TITLE
Homo sapiens chromosome 18, clone RP11-723J4
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 221941)
REFERENCE
AUTHORS
Birn, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,
Cammarata, J., Campoliano, A., Chang, J., Choepel, Y., Colangelo, M.,
Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K.,
Diaz, J., Dodge, S., Faro, S., Ferreira, P., Flizhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W.,
Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
McClean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M.,
McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L.,
Miho, T., Miengo, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
Rella, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
Severy, P., Sougne, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S.,
Theodore, J., Travers, M., Travis, N., Trifilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zaloun, J., Zembek, L., Zimmer, A. and Zody, M.
Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 221941)
REFERENCE
AUTHORS
Birn, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Brown, A., Cammarata, J., Campoliano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J., Dodge, S., Faro, S.,
Ferreira, P., Flizhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Glnde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
McClean, C., MacDonald, P., Marquis, N., Matthews, C.,
McCarthy, M., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Miho, T., Miengo, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Rella, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
```


Koman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schlupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.will.edu

----- Project Information -----

Center project name: L12269

Center clone name: 723 J 4

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846

FEATURES

Location/Qualifiers

Source	Location/Qualifiers
1..221941	/orientation="homologous"
	/db_xref="taxon:9606"
	/chromosome="18"
	/map="18"
	/clone="RP11-723J4"
insure	/clone_lib="RPCJ-11 Human Male BAC"
240..261	
insure	/note="<30 qual SNGL region"
322..326	
insure	/note="<30 qual SNGL region"
635..670	
repeat_region	/rpt_family="(TTTA)n"
repeat_region	complement(674..963)
insure	/rpt_family="Alu5g"
688..715	
insure	/note="<30 qual SNGL region"
951..955	
insure	/note="<30 qual SNGL region"
1010..1013	
repeat_region	/note="<30 qual SNGL region"
1124..1441	
repeat_region	/rpt_family="AluX"
2833..2926	
repeat_region	/rpt_family="(CA)n"
repeat_region	complement(2960..3257)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(3291..3371)
repeat_region	/rpt_family="L2"
3807..4209	
repeat_region	/rpt_family="(LMD)"
repeat_region	complement(4298..4597)
repeat_region	/rpt_family="AluX"
5434..5597	
repeat_region	/rpt_family="FRAM"
repeat_region	complement(6415..6580)
repeat_region	/rpt_family="MIR"
repeat_region	complement(7324..7451)
repeat_region	/rpt_family="MLT1H"
7505..7750	
repeat_region	/rpt_family="AluJb"
repeat_region	complement(7768..7984)
repeat_region	/rpt_family="MLT1H"
8022..8284	
repeat_region	/rpt_family="MLT1J"
repeat_region	complement(8665..9153)
repeat_region	/rpt_family="AluSc"
9584..9680	
repeat_region	/rpt_family="Alu5g/x"
11358..11524	

Query Match	34.7%	Score 787.4	DB 9	Length 221941
Rest Local Similarity	99.9%	Pred. No. 3.7e-184		
Matches 788	Conservative	Mismatches 1	Indels 0	Gaps 0
QY 1	ATGGCGCAGAACTTATCTGTATCTGTGATCTGTGACCTTTGGCCCTCTCTGTGACAAATCCC	60		
Db 187366	ATGGCGCAGAACTTATCTGTATCTGTGATCTGTGACCTTTGGCCCTCTCTGTGACAAATCCC	187307		
QY 61	CTTCATGAACATAAAGCAGCTCTTCCCCAGACCACTGAGAAAATTAAGTCTGAAATTTGG	120		
Db 187306	CTTCATGAACATAAAGCAGCTCTTCCCCAGACCACTGAGAAAATTAAGTCTGAAATTTGG	187247		
QY 121	CAATCTGTGCATTAATGTGTGACTTGGCAATTTCCACACGGCAATATCATCTACAAACACTTT	180		
Db 187246	GAATCTGTGCATTAATGTGTGACTTGGCAATTTCCACACGGCAATATCATCTACAAACACTTT	187187		
QY 181	TTCTACCCGCTATGAGAAAATAATCTTTGTGCTGAGTTGAAGGGTTGAGAAAATTTACTTTCAA	240		
Db 187186	TTCTACCCGCTATGAGAAAATAATCTTTGTGCTGAGTTGAAGGGTTGAGAAAATTTACTTTCAA	187127		
QY 241	AATATAGGCATAGATAAGATTAAAGAAATTAATCTTTGTGCTGAGTTGAAGGGTTGAGAAAATTTACTTTCAA	300		
Db 187126	AATATAGGCATAGATAAGATTAAAGAAATTAATCTTTGTGCTGAGTTGAAGGGTTGAGAAAATTTACTTTCAA	187067		
QY 301	GACCACTGAGCTATCATCTCAGACCATGAGCGTCACTCAGACCATGAGATCATCTCTAGACAC	360		
Db 187066	GACCACTGAGCTATCATCTCAGACCATGAGCGTCACTCAGACCATGAGATCATCTCTAGACAC	187007		
QY 361	GAGCTATCTCTGACCATGATCATCTCTCAGCATTAATCATGCTGCTTCTGTGTAATAAAT	420		
Db 187006	GAGCTATCTCTGACCATGATCATCTCTCAGCATTAATCATGCTGCTTCTGTGTAATAAAT	186947		
QY 421	AAATGAAAAGCTCTTTTCCAGACCATGACCTCAGATGATTCAGTAAAGATCTTAAAC	480		
Db 186946	AAATGAAAAGCTCTTTTCCAGACCATGACCTCAGATGATTCAGTAAAGATCTTAAAC	186887		
QY 481	AGCCAGGGGAAAGAGCTCAGCAGACCAATGATGAGTGGTAGAAGGATGTCAAGGAC	540		
Db 186886	AGCCAGGGGAAAGAGCTCAGCAGACCAATGATGAGTGGTAGAAGGATGTCAAGGAC	186827		
QY 541	AGTCTTACTGCTAGTGAAGTGACCTCAACTGTGTACACACTGTCTCTCAAGAACTTCAC	600		
Db 186826	AGTCTTACTGCTAGTGAAGTGACCTCAACTGTGTACACACTGTCTCTCAAGAACTTCAC	186767		
QY 601	TTTCTAGACACATAGAGACTCCAAAGACCTGGAAGAACTCTTCCCAAGATGTAAATACG	660		
Db 186766	TTTCTAGACACATAGAGACTCCAAAGACCTGGAAGAACTCTTCCCAAGATGTAAATACG	186707		
QY 661	TCCACTTCCACCCACTGTTCATCATCAAGAGCGCGGTGAGCGCGCTGGCTTGTATGAAAAA	720		
Db 186706	TCCACTTCCACCCACTGTTCATCATCAAGAGCGCGGTGAGCGCGCTGGCTTGTATGAAAAA	186647		
QY 721	AATCAATCTGTGACTGAGCCCTCCAAAGGCTTTATGTATTCAGAAACCAATGAAAT	780		
Db 186646	AATCAATCTGTGACTGAGCCCTCCAAAGGCTTTATGTATTCAGAAACCAATGAAAT	186587		
QY 781	CCTCAGGAG 789			
Db 186586	CCTCAGGAG 186578			
RESULT 11				
AP001905				
LOCUS				
DEFINITION	224788 bp DNA linear HTG 11-JUL-2000			
	Homo sapiens chromosome 18 clone RP11-723J4 map 18q12, WORKING			
	DRAFT SEQUENCE, 23 unordered pieces.			
ACCESSION	AP001905			
VERSION	AP001905.3 GI:9188521			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens DNA, clone:RP11-723J4.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			

1 (bases 1 to 224788)	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 224,788 genomic DNA of 18q12	Published Only in Database (2000)
2 (bases 1 to 224788)	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission	Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8117556.	Genome Center
Center: RIKEN Genomic Sciences Center(GSC)	Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/	Contact: hattori@gsc.riken.go.jp
----- Project Information	Center project name: HumDraft18
----- Summary Statistics	Center clone name: RP11-723J4
Sequencing vector: PCR products; 100% of reads	Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329	Consensus quality: 216741 bases at least Q40
Consensus quality: 219708 bases at least Q30	Consensus quality: 221297 bases at least Q20
Insert size: 222586; sum-of-contigs	Quality coverage: 9.33x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved	
1	42444 contig of 42444 bp in length
42545	64659 contig of 22115 bp in length
64760	82876 contig of 18117 bp in length
82977	100906 contig of 17930 bp in length
101007	117951 contig of 16945 bp in length
118052	131503 contig of 13452 bp in length
131604	141537 contig of 9934 bp in length
141638	152889 contig of 11252 bp in length
152990	161818 contig of 8829 bp in length
161919	170983 contig of 9065 bp in length
171084	178239 contig of 7156 bp in length
178340	187572 contig of 9233 bp in length
187673	193032 contig of 5360 bp in length
193133	198842 contig of 5710 bp in length
198943	203032 contig of 4090 bp in length
203133	207793 contig of 4651 bp in length
207884	211428 contig of 3545 bp in length
211529	215591 contig of 4063 bp in length
215692	217832 contig of 2141 bp in length
217933	220510 contig of 2578 bp in length
220611	222035 contig of 1425 bp in length
222136	223655 contig of 1520 bp in length
223756	224788 contig of 1033 bp in length.
* NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
1	42444: contig of 42444 bp in length
42445	42544: gap of 100 bp

* 42545 64659: contig of 22115 bp in length
* 64660 64759: gap of 100 bp
* 64760 82476: contig of 18117 bp in length
* 82477 82976: gap of 100 bp
* 82977 100906: contig of 17930 bp in length
* 100907 101006: gap of 100 bp
* 101007 117951: contig of 16945 bp in length
* 117952 118051: gap of 100 bp
* 118052 131503: contig of 13452 bp in length
* 131504 131603: gap of 100 bp
* 131604 141537: contig of 9934 bp in length
* 141538 141637: gap of 100 bp
* 141638 152889: contig of 11252 bp in length
* 152890 152989: gap of 100 bp
* 152990 161818: contig of 8829 bp in length
* 161819 161918: gap of 100 bp
* 161919 170983: contig of 9065 bp in length
* 170984 171083: gap of 100 bp
* 171084 178239: contig of 7156 bp in length
* 178240 178339: gap of 100 bp
* 178340 187572: contig of 9233 bp in length
* 187573 187672: gap of 100 bp
* 187673 193032: contig of 5360 bp in length
* 193033 193132: gap of 100 bp
* 193133 198842: contig of 5710 bp in length
* 198843 198942: gap of 100 bp
* 198943 203032: contig of 4090 bp in length
* 203033 203132: gap of 100 bp
* 203133 207783: contig of 4651 bp in length
* 207784 207883: gap of 100 bp
* 207884 211428: contig of 3545 bp in length
* 211429 211528: gap of 100 bp
* 211529 215591: contig of 4063 bp in length
* 215592 215691: gap of 100 bp
* 215692 217832: contig of 2141 bp in length
* 217833 217932: gap of 100 bp
* 217933 220510: contig of 2578 bp in length
* 220511 220610: gap of 100 bp
* 220611 222035: contig of 1425 bp in length
* 222036 222135: gap of 100 bp
* 222136 223655: contig of 1520 bp in length
* 223656 223755: gap of 100 bp
* 223756 224788: contig of 1033 bp in length.

FEATURES
 source
 1..224788
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18q12"
 /clone="RP11-723J4"
 1..42444
 /note="assembly_fragment"
42545..64659
 /note="assembly_fragment clone_end:SP6 vector_side:right"
64760..82976
 /note="assembly_fragment"
82977..100906
 /note="assembly_fragment"
101007..117951
 /note="assembly_fragment"
118052..131503
 /note="assembly_fragment"
131604..141537
 /note="assembly_fragment clone_end:W7 vector_side:right"
141638..152889
 /note="assembly_fragment"
152990..161818
 /note="assembly_fragment"
161919..170983
 /note="assembly_fragment"
171084..178239
 /note="assembly_fragment"
178340..187572

misc_feature /note="assembly_fragment"
187673..193032
misc_feature /note="assembly_fragment"
193133..198842
misc_feature /note="assembly_fragment"
198943..203032
misc_feature /note="assembly_fragment"
203133..207783
misc_feature /note="assembly_fragment"
207884..211428
misc_feature /note="assembly_fragment"
211529..215591
misc_feature /note="assembly_fragment"
215692..217832
misc_feature /note="assembly_fragment"
217933..220510
misc_feature /note="assembly_fragment"
220611..222035
misc_feature /note="assembly_fragment"
222136..223655
misc_feature /note="assembly_fragment"
223756..224788
/note="assembly_fragment"
BASE COUNT 67462 a 44418 c 45111 g 65596 t 2201 others
ORIGIN

Query Match 34.7% Score 787.4; DB 2; Length 224788;
Best Local Similarity 99.9%; Pred. No. 3.7e-184;
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCGGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTGGCCCTCTCTGTGCACAAATCCC 60
 |||||
Db 2349 ATGCGGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTGGCCCTCTCTGTGCACAAATCCC 2408

Qy 61 CTTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACCACCTGAGAAATTAAGTCCGAATTGG 120
 |||||
Db 2409 CTTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACCACCTGAGAAATTAAGTCCGAATTGG 2468

Qy 121 GAATCTGGCATTATGTTGACTTGGCAATTTCCACAGCGGAATATCATCTACACACGCTT 180
 |||||
Db 2469 GAATCTGGCATTATGTTGACTTGGCAATTTCCACAGCGGAATATCATCTACACACGCTT 2528

Qy 181 TTCTACCGCTATGAGAAATAATTTCTGTGACCTTGAAGGGTTTCAGAAAATTAAGTCTTCAA 240
 |||||
Db 2529 TTCTACCGCTATGAGAAATAATTTCTGTGACCTTGAAGGGTTTCAGAAAATTAAGTCTTCAA 2588

Qy 241 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCACGACCATCACTCA 300
 |||||
Db 2589 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCACGACCATCACTCA 2648

Qy 301 GACCACGAGCATCACTCAGACCACCTGAGCGTCACTCAGACCACCTGAGCATCACTCAGACCAC 360
 |||||
Db 2649 GACCACGAGCATCACTCAGACCACCTGAGCGTCACTCAGACCACCTGAGCATCACTCAGACCAC 2708

Qy 361 GAGCATCACTCTGACCATGATCATCACTCTCAGCAATAATCATGCTGCTTCTGGTAAAAAT 420
 |||||
Db 2709 GAGCATCACTCTGACCATGATCATCACTCTCAGCAATAATCATGCTGCTTCTGGTAAAAAT 2768

Qy 421 AAGCGAAAAGCTCTTTGCCCGAGACCACCTGAGTTCAGTAAAGATCCTAGAAAC 480
 |||||
Db 2769 AAGCGAAAAGCTCTTTGCCCGAGACCACCTGAGTTCAGTAAAGATCCTAGAAAC 2828

Qy 481 AGCCAGGGGAAAGAGCTCAACCGACCAAGATGCCAGTGGTAGAAGGATGTCAAGGAC 540
 |||||
Db 2829 AGCCAGGGGAAAGAGCTCAACCGACCAAGATGCCAGTGGTAGAAGGATGTCAAGGAC 2888

Qy 541 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGGAACCTCAC 600
 |||||
Db 2889 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGGAACCTCAC 2948

Qy 601 TTTTCTAGACAAATAGAGACTCCCAAGACCTGGAAAACTCTTCCCCAAAGATGTGAAGCAGC 660
 |||||
Db 2949 TTTTCTAGACAAATAGAGACTCCCAAGACCTGGAAAACTCTTCCCCAAAGATGTGAAGCAGC 3008

QY	661	TCCACTTCACCGCAGTGTGCATCTCAAAAGAGCCGGTGAGCCGCGTGGCTGGTAGGAAAACA	720
Df	3009	TCCACTTCACCGCAGTGTGCATCTCAAAAGAGCCGGTGAGCCGCGTGGTAGGAAAACA	1068
QY	721	AATGAATCTGTGACTGACGCCGAAAAAGCTTTATGTATTCCAGANAACACAAATCAAAT	780
Df	3069	AATGAATCTGTGACTGACGCCGAAAAAGCTTTATGTATTCCAGANAACACAAATCAAAT	1128
QY	781	CCTCAGGAG 789	
Df	3129	CCTCAGGAG 3137	
RESULT 12			
AX465590		AX465590	899 bp
LOCUS		Sequence 3 from Patent WO0216939.	
DEFINITION		AX465590	
ACCESSION		AX465590.1	GI:21899893
VERSION			
KEYWORDS		Mus sp.	
SOURCE		Mus sp.	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE		1 Mammalia; Euheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		Mack,D., Gish,K.C. and Wilson,K.E.	
TITLE		Methods of diagnosis of cancer and screening for cancer modulators	
JOURNAL		Patent: WO 0216939-A 3 28-FEB-2002;	
FEATURES		EOS Biotechnology, Inc. (US)	
SOURCE		location/Qualifiers	
		1..899	
		/organism="Mus Sp."	
		/db_xref="taxon:10095"	
		/note="prostate/breast/ovarian/bladder cancer protein	
		mouse BCR4 cDNA"	
		4...>897	
		/note="mouse BCR4"	
		/codon_start=1	
		/protein_id="CAD42375.1"	
		/db_xref="GI:21899894"	
		/translation="MSIQALNNATEFNYPALINIDAKNLIHTASEKKAEIPPKT	
		YSLSIAWLGGFIAISIFSLISLIGVILVPLMNRVPFKFLISFLVALAVGTLSGDALLH	
		LILPHSHASHQHISHHEPAMEMKRGPLTHISAONTSEESYSTSTWKGLTAIGGLGYFM	
		FYLRHVLTLLIKOFKDKKXXKKXPPENDEVESKKOLSKYDLOSLSNEEKVDPERPES	
		VLRADSOEPSDPDSQQDDVGRCRHXSDSPCTPHKKSTMNVPRGIEQEPFTILSRVAG	
		PVRPRXHIIHDVHHIIXHLTR"	
HASK COUNT	252 a	239 c	198 g
ORIGIN			203 t
			7 others
Query Match		27.3%;	Score 619.6; DB 6;
Best local Similarity		84.3%;	Pred. No. 8.9e-143;
Matches	752; Conservative	1;	Mismatches 131; Indels 8; Gaps 5;
QY	823	GCGATGGGCATCCAGGTTCCCGCTGAATGCAACAGAGTTCAACTATCTCTCGAGCCATC	882
Df	1	GCGATGAGCATCCAGGCTCTGTTGAATGCAACAGCAATTTAATCTATCTCTCCAGCATC	60
QY	883	ATCAACCAAAATTCATGCTAGATCTTGTCTGATTTCATA--CAAGTGAAGAAGAGCTGAA	939
Df	61	ATCAATCAAAATTCATGCTCGGGCTGTGCTGATTCATACAGCAATGAGAGAAGAGGAGAA	120
QY	940	ATCCCTCCAAGACCTATTTCATTACAAATAGCCCTGGCTTGGTGGTTTATAGCCATTTC	999
Df	121	ATCCCTCCAAGACCTATTTCATTACAAATAGCCCTGGCTTGGTGGCTTTCATAGCCATTTC	180
QY	1000	ATCATCAGTTTTCTGTCTGTCTGTCTGGGGGTATTCTTATAGCCCTCTCATGAATCGGCTTT	1059
Df	181	ATCATCAGTTTTCTGTCTGTCTGTCTGGGAGTCATCTTGTGTCACCTCATGAACCGGCTATT	240
QY	1060	TTCAAAATTTCTCTGAGTTTCTCTTGTGGCACCTGGCGCTTGGGACTTTGACTGATGATCT	1119
Df	241	TTCAAGTTCTCTGAGCTTCTGTCTGTCTGGGGCTGGCCCTGGAACCGCTGAGTGTGCTATGCT	300

Oraqunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, K., Primus, E., Pu, L.L., Quiles, M., Ken, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshkari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sulton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tanq, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villafion, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 193437)
 Worley, K.C.
 Direct Submission
 Submitted (08-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193437)
 Worley, K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:21703549.

 Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: CH230-5p17
 Center clone name: CH230-5p17

 Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 143517 bases at least Q40
 Consensus quality: 150307 bases at least Q30
 Consensus quality: 154257 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1172: contig of 1172 bp in length
 * 1173 1272: gap of unknown length
 * 1273 1272: gap of unknown length
 * 2488 2487: contig of 1215 bp in length
 * 2588 2587: gap of unknown length
 * 4395 4394: contig of 1807 bp in length
 * 4495 4494: gap of unknown length
 * 5816 5815: contig of 1321 bp in length
 * 5916 5915: gap of unknown length
 * 7627 7626: contig of 1711 bp in length
 * 7727 7726: gap of unknown length
 * 9207 9206: contig of 1480 bp in length
 * 9307 9306: gap of unknown length
 * 10450 10449: contig of 1143 bp in length
 * 10550 10549: gap of unknown length
 * 11610 11609: contig of 1060 bp in length
 * 11710 11709: gap of unknown length
 * 12958 12957: contig of 1248 bp in length
 * 13058 13057: gap of unknown length
 * 14502 14501: contig of 1444 bp in length
 * 14601: gap of unknown length
 *
 * 14602 15762: contig of 1161 bp in length
 * 15762: gap of unknown length
 * 15862 15862: contig of 1015 bp in length
 * 16877 16877: gap of unknown length
 * 16978 16977: gap of unknown length
 * 18761 18761: contig of 1784 bp in length
 * 18861 18861: gap of unknown length
 * 20537 20537: contig of 1676 bp in length
 * 20538 20537: gap of unknown length
 * 20638 20638: contig of 2093 bp in length
 * 22731 22730: gap of unknown length
 * 22831 22830: gap of unknown length
 * 25262 25262: contig of 2432 bp in length
 * 25362 25362: gap of unknown length
 * 25363 25363: contig of 1900 bp in length
 * 27263 27262: gap of unknown length
 * 27363 27362: contig of 1162 bp in length
 * 28525 28524: gap of unknown length
 * 28625 28624: gap of unknown length
 * 30233 30233: contig of 1609 bp in length
 * 30334 30333: gap of unknown length
 * 31884 31883: contig of 1550 bp in length
 * 31984 31983: gap of unknown length
 * 33437 33436: contig of 1453 bp in length
 * 33537 33536: gap of unknown length
 * 34740 34739: contig of 1203 bp in length
 * 34840 34839: gap of unknown length
 * 36918 36917: contig of 2078 bp in length
 * 37018 37017: gap of unknown length
 * 38174 38173: contig of 1156 bp in length
 * 38274 38273: gap of unknown length
 * 39953 39952: contig of 1679 bp in length
 * 40053 40052: gap of unknown length
 * 42001 42000: contig of 1948 bp in length
 * 42101 42100: gap of unknown length
 * 43703 43702: contig of 1602 bp in length
 * 43803 43802: gap of unknown length
 * 46150 46149: contig of 2347 bp in length
 * 46250 46249: gap of unknown length
 * 47599 47598: contig of 1349 bp in length
 * 47659 47658: gap of unknown length
 * 49211 49210: contig of 1512 bp in length
 * 49311 49310: gap of unknown length
 * 50781 50780: contig of 1470 bp in length
 * 50881 50880: gap of unknown length
 * 52851 52850: contig of 1970 bp in length
 * 52951 52950: gap of unknown length
 * 53981 53980: contig of 1031 bp in length
 * 54082 54081: gap of unknown length
 * 55921 55920: contig of 1740 bp in length
 * 57909 57908: gap of unknown length
 * 58009 58008: contig of 1988 bp in length
 * 60388 60387: gap of unknown length
 * 60489 60488: contig of 2379 bp in length
 * 62461 62460: gap of unknown length
 * 62561 62560: gap of unknown length
 * 64501 64500: contig of 1940 bp in length
 * 66890 66889: gap of unknown length
 * 66990 66989: gap of unknown length
 * 69590 69589: gap of unknown length
 * 71654 71653: contig of 1964 bp in length
 * 71754 71753: gap of unknown length
 * 75274 75273: contig of 3520 bp in length
 * 75374 75373: gap of unknown length
 * 77773 77772: contig of 2399 bp in length
 * 77873 77872: gap of unknown length
 * 81414 81413: contig of 3541 bp in length
 * 81514 81513: gap of unknown length
 * 85250 85249: contig of 3736 bp in length
 * 85350 85349: gap of unknown length
 * 87367 87366: contig of 2017 bp in length
 * 87467 87466: gap of unknown length
 * 90125: contig of 2659 bp in length

```
* 90126 90225: gap of unknown length
* 90226 93089: contig of 2864 bp in length
* 93189 93190: gap of unknown length
* 93190 96456: contig of 3267 bp in length
* 96457 96556: gap of unknown length
* 96557 99751: contig of 3195 bp in length
* 99752 99851: gap of unknown length
* 99852 103352: contig of 3501 bp in length
* 103353 103453: gap of unknown length
* 103453 107023: contig of 3571 bp in length
* 107024 107123: gap of unknown length
* 107124 112368: contig of 5245 bp in length
* 112369 112469: gap of unknown length
* 112469 115712: contig of 3244 bp in length

Query Match 18.6% Score 421.4; DB 2; Length 19,437;
Rest Local Similarity 75.7%; Pred. No. 2.6e-93;
Matches 606; Conservative 0; Mismatches 126; Indels 69; Gaps 4;

QY 1 ATGGCGAGGAAGTTATCTGTAATCTTGATCTTGACCTTTGGCCCTCTCTGTGCACAAATGCC 60
Db 107530 ATGGCCACAAATTTATCTGTAATCATGATCTTGACCTTTGGCCCTTTGGGTTACAAATCCC 107589

QY 61 CTTTCATGAACATAA---AGCAGCTGCTTTCCCCAGACCACTGAGAAAATTAGTCCGAAT 117
Db 107590 CTTTCATGAACATAAATCAACAGCTGCTTTCTCTCAGACCACTGAGAAAATTAATTCAAAT 107649

QY 118 TCGGAATCTGGCATTTAATCTTGACCTTGCCCAATTTCCACACGCCAATATCATCTACAACAG 177
Db 107650 TCGGAATCTGGCATTTAATCTTGACCTTGCCCAATTTCCACACGCCAATATCATCTACAACAG 107709

QY 178 CTTTTCATGCGGTATGGAGAAATAATTTTGTGAGTTGAAGGTTTCAGAAAATTTACTT 237
Db 107710 CTTTTCATGCGGTATGGAGAAATAATTTTGTGAGTTGAAGGTTTCAGAAAATTTACTT 107769

QY 238 CAAATATAGGCATAGATAAGATTAAGAATCCATATACACCATGACCCAGCCATACAC 297
Db 107770 CAAATATAGGCATAGATAAGATTAAGAATCCATATACACCATGACCCAGCCATACAC 107820

QY 298 TCAGACACGAGCATCACTCAGACCATGAGCGTCACCTCAGACCATGAGCATCTCAGAC 357
Db 107821 TCAGACACGAGCATCACTCAGACCATGAGCGTCACCTCAGACCATGAGCATCTCAGAC 107835

QY 358 CACGAGCATCACTGACCATGATCATCACTCTCACCATAATCATGCTGCTTCTGTAATA 417
Db 107836 CACGAGCATCACTGACCATGATCATCACTCTCACCATAATCATGCTGCTTCTGTAATA 107895

QY 418 AATAAGCCAAAAGCTCTTTGGCCAGACCATCACTCAGATAGTTTCAGGTAAGATCTTAGA 477
Db 107896 AATAAGCCAAAAGCTCTTTGGCCAGACCATCACTCAGATAGTTTCAGGTAAGATCTTAGA 107955

QY 478 AACAGCTCAGGGGAAGGAGCTCAGCCAGACAGATGCCAGTGGTAGAAGGAATGTCAG 537
Db 107956 AACAGCTCAGGGGAAGGAGCTCAGCCAGACAGATGCCAGTGGTAGAAGGAATGTCAG 108015

QY 538 GACAGCTGTTAGTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 597
Db 108016 GACAGCTGTTAGTGTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 108075

QY 598 CACTTCTTAGAGACAAATAGACATCCAGACCTGGAAGACTCTTCCCAAGAGATCTAAGC 657
Db 108076 CACTTCTTAGAGACAAATAGACATCCAGACCTGGAAGACTCTTCCCAAGAGATCTAAGC 108132

QY 658 AGCTCCACTCCACCCAGTGTACATCAAAAGAGCGGGTGGCCCGGCTG-----GCT 708
Db 108133 AGCTCCACTCCACCCAGTGTACATCAAAAGAGCGGGTGGCCCGGCTG-----GCT 108192

QY 709 GGTAGGAACAAATGATCTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 768
Db 108193 GGTAGGAACAAATGATCTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 108252

QY 769 ACAAATGAAATCTCTCAGGAG 789
Db 108252 ACAAATGAAATCTCTCAGGAG 789
```

Db 108253 TCAATGACATATATCCAGGAG 108273

Search completed: November 28, 2002, 04:15:24
Job time : 6214 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 27, 2002, 23:26:15 ; Search Time 2851 Seconds

(without alignments)

12883.677 Million cell updates/sec

Title: US-09-642-034-4

Perfect score: 2268

Sequence: 1 atggcaggaagtattctgt.....tglttcyatalaaatttctay 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba: *

2: em_esthm: *

3: em_estin: *

4: em_estnu: *

5: em_estov: *

6: em_estpl: *

7: em_estro: *

8: em_hlc: *

9: gb_est1: *

10: gb_est2: *

11: gb_hlc: *

12: gb_est3: *

13: gb_est4: *

14: gb_est5: *

15: em_estfun: *

16: em_estom: *

17: gb_gss: *

18: em_gss_hum: *

19: em_gss_inv: *

20: em_gss_pln: *

21: em_gss_vrt: *

22: em_gss_fun: *

23: em_gss_mam: *

24: em_gss_mus: *

25: em_gss_other: *

26: em_gss_pro: *

27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	877.6	38.7	1074	13	BM480018
2	799.6	35.3	872	9	BM480018
3	724	31.9	810	13	BM456317
4	692	30.5	701	14	BM456317
5	652	28.7	1135	13	BM560789
6	649.8	28.7	674	14	BM978669

7	627.6	27.7	957	12	BG168169
8	600.2	26.5	746	12	BQ530601
9	597.6	26.3	604	14	BQ581865
10	586.8	25.9	766	14	BM948026
11	576	25.4	925	12	BF983458
12	571.4	25.2	743	9	AI907176
13	567.4	25.0	829	9	AU079709
14	566.8	25.0	751	14	BM963737
15	566.2	25.0	887	13	BI150324
16	564	24.9	781	14	BM949686
17	563.8	24.9	857	12	BF032013
18	538.6	23.7	948	14	BO955836
19	526.6	23.2	580	12	BF915495
20	517.8	22.8	849	12	BF680209
21	515.8	22.7	791	12	BG547794
22	505.2	22.3	682	14	BM949679
23	503.4	22.2	950	14	BQ433766
24	502.2	22.1	518	10	AW178411
25	501.8	22.1	513	10	AW178409
26	499.6	22.0	561	9	AL042316
27	499.2	22.0	1003	13	BM478134
28	495.8	21.9	560	12	BF915709
29	493.6	21.8	500	10	AW178317
30	493.4	21.8	495	10	AW178381
31	493.4	21.8	522	10	AW383696
32	493.4	21.8	522	10	AW383720
33	489.6	21.6	518	10	AW383693
34	489.6	21.6	518	10	AW383717
35	481.6	21.2	1112	14	BQ932626
36	475	20.9	593	10	AW365994
37	474	20.9	933	13	BI861563
38	472.4	20.8	507	12	BF699194
39	472.2	20.8	739	9	AA666965
40	469	20.7	694	9	AA666964
41	459.8	20.3	605	13	BI284823
42	458	20.2	1059	13	BM557996
43	454	20.0	528	9	AA242758
44	453.2	20.0	501	10	AW994833
45	446	19.7	484	13	BG994120

ALIGNMENTS

RESULT 1	BM480018	BM480018	1074 bp	mRNA	linear	EST 05-FEB-2002
LOCUS	AGENCOURT_6424145	NIH_MGC_67	Homo sapiens	cdna	clone IMAGE:5491572	
DEFINITION	5', mRNA sequence.					
ACCESSION	BM480018					
VERSION	BM480018.1	GI:18529060				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 1074)					
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.					
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)					
	Unpublished (1997)					
	Contact: Robert Strausberg, Ph.D.					
	Email: cgapbs-r@mail.nih.gov					
	Tissue Procurement: ATCC					
	CDNA Library Preparation: Life Technologies, Inc.					
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)					
	DNA Sequencing by: Agencourt Bioscience Corporation					
	Clone distribution: MGC clone distribution information can be					
	found through the I.M.A.G.E. Consortium/LLNL at:					
	http://image.llnl.gov					
	Plate: LLAM12112 row: e column: 13					
	High quality sequence stop: 652.					
	Location/Qualifiers					
	1. .1074					

FEATURES
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5491572"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="pR10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SpQMT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
326 a 256 c 220 g 271 t 1 others
BASE COUNT
ORIGIN

```

Query Match	38.7%	Score 877.6	DB 13	Length 1074
Best Local Similarity	94.4%	Pred. No. 4.3e-186		
Matches 954	Conservative 0	Mismatches 50	Indels 7	Gaps 4
Qy 416	AAATAAGCGAAAAGCTCTTTTGCCACAGACCATGACTCAGATGATTCAGTAAAGATCCTA	475		
Db				
4	AAATAAGCGAAAAGCTCTTTTGCCACAGACCATGACTCAGATGATTCAGTAAAGATCCTA	63		
Qy 476	GAACAGCGCCAGGGGAAAGGAGCTCACCGCCACAGAACATGCCAGTGGTATAGAGGAATTCGA	535		
Db				
64	CAAAACGCCAGGGGAAAGGAGCTCACCGCCACAGAACATGCCAGTGGTATAGAGGAATTCGA	123		
Qy 536	AGGACACTGTTAGTTCCTACTCAAGTGGACCTCAACTGTGTACAACACTGTCTCTGAAGAA	595		
Db				
124	AGGACACTGTTAGTTCCTACTCAAGTGGACCTCAACTGTGTACAACACTGTCTCTGAAGAA	183		
Qy 596	CTCACTTTTCTAGACACAATAGACACTCCAAAGACTCGGAAACCTCTTCCCAAGACATCTAA	655		
Db				
184	CTCACTTTTCTAGACACAATAGACACTCCAAAGACTCGGAAACCTCTTCCCAAGACATCTAA	243		
Qy 656	GCACCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGCTGGCTGGTAGGA	715		
Db				
244	GCAGCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGCTGGCTGGTAGGA	303		
Qy 716	AAACAATAATGTAATCTGTGAGTGGCCCCCAAGGCTTTATGTATTCAGAAACACAATG	775		
Db				
304	AAACAATAATGTAATCTGTGAGTGGCCCCCAAGGCTTTATGTATTCAGAAACACAATG	363		
Qy 776	AAATCTCTCAGGAGTGTTCATATGCATCAAGCTACATGACATCTCATGGCATGGCATCC	835		
Db				
364	AAATCTCTCAGGAGTGTTCATATGCATCAAGCTACATGACATCTCATGGCATGGCATCC	423		
Qy 836	AGTTTCCGCTCAATGTCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAAGCTCAATG	895		
Db				
424	AGTTTCCGCTCAATGTCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAAGCTCAATG	483		
Qy 896	ATGCTAGATCTTGCTGATTCATACAAGTGAAGAGAGGCTGAATCCCTCCAAAGACCT	955		
Db				
484	ATGCTAGATCTTGCTGATTCATACAAGTGAAGAGAGGCTGAATCCCTCCAAAGACCT	543		
Qy 956	ATTTCATTACAAATAGCGTGGTGGTGTATAGCCATTCCTCATCATCAGTTCCTGT	1015		
Db				
544	ATTTCATTACAAATAGCGTGGTGGTGTATAGCCATTCCTCATCATCAGTTCCTGT	603		
Qy 1016	CTCTGCTGGGGTTATCTTAGTGCCCTCTCATGAATCGGGTGTTCATATTCCTCTGA	1075		
Db				
604	CTCTGCTGGGGTTATCTTAGTGCCCTCTCATGAATCGGGTGTTCATATTCCTCTGA	663		
Qy 1076	GTTCCTCTTGGGCACGTGGCGTTGGGACATTGAGTGGTGATGCTTTTACACCTTCCTTC	1135		
Db				
664	GTTCCTCTTGGGCACGTGGCGTTGGGACATTGAGTGGTGATGCTTTTACACCTTCCTTC	723		
Qy 1136	CACATTTCTCATGCCAAGTCAACCACCATAGTCATAGCCATGAAGAACCCAGCAATCGAAATGA	1195		
Db				
724	CACATTTCTCATGCCAAGTCAACCACCATAGTCATAGCCATGAAGAACCCAGCAATGA	783		
Qy 1196	AAAGAGGACCACTTTTTCAGTCACTGTCTTCTCAAAACATAGAAGAAAGTGGCTATTG	1255		
Db				
784	AAAGAGGACCACTTTTTCAGTCACTGTCTTCTCAAAACATAGAAGAAAGTGGCTATTG	842		

QY	1256	ATTCCACGTTGAAG-IGCTAAACAGCTCTAGGAGCGCTGATTTCATGTTCTCTGTTGAAC	1315
DB	843	ATTTCCCGTGGAA-GGCTTAACAGCTCCCTAGAAGCCCGCATTTCTGTTTCTCTGTTGAAC	901
QY	1316	ATGTCTCATCATGATCAACAACATTTAAAGATTAAGAGAGAAAGAAATCAGAGAAACCTG	1375
DB	902	ATGCCCGCCCAATGGTGC-AACCCCTTTTGGGTAGGAAGAAAGAAATCCCTCAAAAAACCTG	960
QY	1376	AAA-----ATGATGATGATGTCGAGATTTAAGAAGCAGTGTCTCCCAAGTATGAA	1422
DB	961	AAAAATGCATGATGATGTTGGAATTTAAAAAACCGTTTGCCCGCCCAATGAA	1011
RESULT 2			
AUI20027			
LOCUS			
DEFINITION	AUI20027	HEMBAL Homo sapiens cDNA clone HEMBA1007185 5', mRNA	EST 01-AUG-2007
ACCESSION			
VERSION	AUI20027		
KEYWORDS	AUI20027.1	G1:10935262	
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			

[illegible]


```
QY 1536 GGAAGAGAGAGGTCATGATGATGCTCATGCTCATCCACAGGAAGTCTACAAATTAATATGT 1595
DB 1536 GGAAGAGAGAGGTCATGATGATGCTCATGCTCATCCACAGGAAGTCTACAAATTAATATGT 300
QY 1596 ACCAGAGAGGTCGAAGAAATAAATGCCATTTACATTTCCAGATACACATCGGCCAGTCCAGA 1655
DB 1596 ACCAGAGAGGTCGAAGAAATAAATGCCATTTACATTTCCAGATACACATCGGCCAGTCCAGA 360
QY 1656 CGATCTCATTTCCACCACCATGATGACTACCATCATATTTCCATCATCACACACCAACAAA 1715
DB 1656 CGATCTCATTTCCACCACCATGATGACTACCATCATATTTCCATCATCACACACCAACAAA 420
QY 1716 CCATCATCTCTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTGAAGATGCGCGCT 1775
DB 1716 CCATCATCTCTCACAGTCACAGCCAGCGCTACTCTCGGAGGAGCTGAAGATGCGCGCT 480
QY 1776 CGCCACATTTGGCCCTGGATGATGATAATGGGTGATGGCTGCACAAATTTCAAGCATGCGCT 1835
DB 1776 CGCCACATTTGGCCCTGGATGATGATAATGGGTGATGGCTGCACAAATTTCAAGCATGCGCT 540
QY 1836 AGCAATTTGCTGCTGCTTTTACTGAAGCTTATCAAGTGGTTTAAAGTACTTCTGTTGCTGT 1895
DB 1836 AGCAATTTGCTGCTGCTTTTACTGAAGCTTATCAAGTGGTTTAAAGTACTTCTGTTGCTGT 600
QY 1896 GTTCTGTCATGAGTTGGCTCTCATTAATTTAGGTGACTTTTCTCTACTAAAGGCTGGCAT 1955
DB 1896 GTTCTGTCATGAGTTGGCTCTCATTAATTTAGGTGACTTTTCTCTACTAAAGGCTGGCAT 660
QY 1956 GACCTTTAAAGCAGGCTGTCTTTTATAATGATGATGATGATGATGATGATGATGATGATG 2015
DB 1956 GACCTTTAAAGCAGGCTGTCTTTTATAATGATGATGATGATGATGATGATGATGATGATG 720
QY 2016 GCAACAGGAATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2075
DB 2016 GCAACAGGAATTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 2076 TACTGCTGGCTTATTCATGATGATGTTG--CTCTGCTGATATGTTGATGATGATGATGATG 2133
DB 2076 TACTGCTGGCTTATTCATGATGATGTTG--CTCTGCTGATATGTTGATGATGATGATGATG 840
QY 2134 AAT--GATGCTAGTACCATGG 2153
DB 2134 AATGATGCTAGTACCATGG 861

RESULT 3
LOCUS BM456317 810 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6409853 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5496678
5', mRNA sequence.
ACCESSION BM456317
VERSION BM456317.1 GI:18505357
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 810)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cqaabs-remail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12125 row: j column: 07
High quality sequence stop: 640.
```

```
FEATURES
Source
Location/Qualifiers
1..810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5496678"
/culture_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
1 others
BASE COUNT 246 a 172 c 165 g 226 t
ORIGIN
Query Match 31.9%; Score 724; DB 13; Length 810;
Best Local Similarity 96.2%; Pred. No. 1.1e-151;
Matches 764; Conservative 0; Mismatches 25; Indels 5; Gaps 2;
QY 866 ATCTCTGTCAGCCATCATCAACCAATTTGATGCTAGATCTTGTCTGTGATTCATACAAGTG 925
DB 1 ATCTCTGTCAGCCATCATCAACCAATTTGATGCTAGATCTTGTCTGTGATTCATACAAGTG 60
QY 926 AAAAGAGGCTGAAATCCCTCCAAAGACCTATTTCATTACAAATAGCCTGGGTGGTGGT 985
DB 61 AAAAGAGGCTGAAATCCCTCCAAAGACCTATTTCATTACAAATAGCCTGGGTGGTGGT 120
QY 986 TTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1045
DB 121 TTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 180
QY 1046 TGAATCGGTGTTTTCAAATTTCTCCTCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1105
DB 181 TGAATCGGTGTTTTCAAATTTCTCCTCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
QY 1106 TGAGTGGTGTATGCTTTTACACCTTTCTTCCACATTTCTCCTCAAGTCAACCAATAGTC 1165
DB 241 TGAGTGGTGTATGCTTTTACACCTTTCTTCCACATTTCTCCTCAAGTCAACCAATAGTC 300
QY 1166 ATAGCCATGAAGACAGCAATGGAAATGAAAGAGGACCACTTTTTCAGTCACTCTCTCT 1225
DB 301 ATAGCCATGAAGACAGCAATGGAAATGAAAGAGGACCACTTTTTCAGTCACTCTCTCT 360
QY 1226 CTCAAACATAGAAGAAAGTGCCTATTTTGATTCACGCTGGAGGGTCTAACAGCTCTAG 1285
DB 361 CTCAAACATAGAAGAAAGTGCCTATTTTGATTCACGCTGGAGGGTCTAACAGCTCTAG 420
QY 1286 GAGGCTGTATTTTCATGTTTCTTGTGTAACATGTCTCTCACTTGTATCAACCAATTTTAAAG 1345
DB 421 GAGGCTGTATTTTCATGTTTCTTGTGTAACATGTCTCTCACTTGTATCAACCAATTTTAAAG 480
QY 1346 ATAAGAAAGAAAGATCAGAGAACCTTCAACCAATGAGAGAAAGTAGATACAGATGATC 1405
DB 481 ATAAGAAAGAAAGATCAGAGAACCTTCAACCAATGAGAGAAAGTAGATACAGATGATC 540
QY 1406 AGTTGTCCAAAGTATGAATCTCAACTTTCAACCAATGAGAGAAAGTAGATACAGATGATC 1465
DB 541 AGTTGTCCAAAGTATGAATCTCAACTTTCAACCAATGAGAGAAAGTAGATACAGATGATC 600
QY 1466 GAACCTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCACCTTTTGTATCTCAGCAGC 1525
DB 601 GAACCTGAAGGCTATTATACGAGCAGACTCACAGAGCCCTCCACCTTTTGTATCTCAGCAGC 660
QY 1526 CTGCGAGTCTTGAAGAAAGAGGTCATGATAGCTCATGCTCATCCACAGG--AAGTCTAC 1584
DB 661 CTGCGAGTCTTGAAGAAAGAGGTCATGATAGCTCATGCTCATCCACAGGCAAGTCTAC 720
QY 1585 AATGAATATGTACCC-----AGAGGCTGCAAGAAATAAATGCCATTTCACATTTCCACGATAC 1640
DB 721 AATGAATATGTAAACCCAGAGAGGTCGCCAAGAAATAAATGCCATTTCACATTTTCACCGAT 780
QY 1641 ACTCGGCCAGTCAG 1654
```

```

Db      781 ACAC"TTCCGGCCAG 794
|||||
H0027619 701 bp mRNA linear EST 27-MAR-2002
UI-H-COO-aq2-a-11-0-01.s1 NCL_CGAP_Sub9 Homo sapiens cDNA clone
IMAGE:3105668 3', mRNA sequence.
H0027619 80027619
H0027619.1 GI:19762898
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 701)
NCL_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Seq primer: M3 FORWARD
POL:YA-Yes.
Location/Qualifiers
1..701
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3105668"
/clone_lib="NCL_CGAP_Sub9"
/lisuse_type="mixed"
/lab_stage="mixed"
/lab_host="DH10B (Life Technologies)"
note="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker; Site1: EcoR I; Site2: Not I; Lissues:
Cholonic mucosa with Crohns disease, Cholonic mucosa with
ulcerative colitis, Fetal thymus, Cervix, Cervical
adenosquamous carcinoma, Ligament cells, Prostate
carcinoma, Bladder carcinoma, Brain oligodendrocyte ;
NCL_CGAP_Sub9 is a substracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are CGTC, AAGC, GGCC,
GGAAG, TAGC, TAAGC, ATGG, AGACA, ATCAC. For additional
information, contact: Bento Soares, bento-soares@uiowa.edu
TAG_LIB=UI-H-COO
TAG_TISSUE=Prostate Carcinoma
TAG_SEQ=ATGC"
H0027619 198 a 152 g 208 t
BASE COUNT 198 a 152 g 208 t
ORIGIN
Query Match 30.5%; Score 692; DB 14; Length 701;
Best local Similarity 99.3%; Pred. No. 1.6e-144;
Matches 695; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 668 CACCCAGTGTACATCAAGAGCCGGCTGAGCGGCTGGTGTAGGAAAAAATAAT 727
|||||
Db 701 CACCCAGTGTACATCAAGAGCCGGCTGAGCGGCTGGTGTAGGAAAAAATAAT 642
QY 728 CTGTGAGTGAGCCCGGAAAGGCTTTATGTATTCAGAGAACACAAATGAAATTCCTAG 787

```

FEATURES
source

```

Db      641 CTGTGAGTGAGCCCGAAAGGCTTTATGTATTCAGAGAACACAAATGAAATCCTCAGG 582
QY 788 AGTGTTCATATGCATCAAGCTACTTGACATCTCATGGCATGGGCATCCAGGTTCCGCTGA 847
|||||
Db 581 AGTGTTCATATGCATCAAGCTACTTGACATCTCATGGCATGGGCATCCAGGTTCCGCTGA 522
QY 848 ATGCAGACAGAGTTCAACTATCTCTCTCAGCCATCATCAACCAAAATTCATGCTAGATCTT 907
|||||
Db 521 ATGCAGACAGAGTTCAACTATCTCTCTCAGCCATCATCAACCAAAATTCATGCTAGATCTT 462
QY 908 GTCTGATTCATACAGTCAAGAGGCTGAAATCCCTCCCAAGACCTTATTCATTACAAA 967
|||||
Db 461 GTCTGATTCATACAGTCAAGAGGCTGAAATCCCTCCCAAGACCTTATTCATTACAAA 402
QY 968 TAGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1027
Db 401 TAGCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 342
QY 1028 TTATCTTAGTGCCTCTCATGAATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1087
|||||
Db 341 TTATCTTAGTGCCTCTCATGAATCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 282
QY 1088 CACTGGCGCTGGGACTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1147
|||||
Db 281 CACTGGCGCTGGGACTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 222
QY 1148 CAAGTCACCCATAGTATAGCCATGAAGAACAGCAGCAATGGAATGAAAGAGAGGACCAC 1207
|||||
Db 221 CAAGTCACCCATAGTATAGCCATGAAGAACAGCAGCAATGGAATGAAAGAGAGGACCAC 162
QY 1208 TTTTCAGTCATCTGTCTCTCAAAACATAGAGAAAGTGCCTATTTTTCATTCACATCTCATG 1267
|||||
Db 161 TTTTGTAGTCATCTGTCTCTCAAAACATAGAGAAAGTGCCTATTTTTCATTCACATCTCATG 102
QY 1268 AGGCTCTAACAGCTCTAGGAGGCTGTATTTTCATGTTTGTGAACATGCTCCTCACAT 1327
Db 101 AGGCTCTAACAGCTCTAGGAGGCTGTATTTTCATGTTTGTGAACATGCTCCTCACAT 42
QY 1328 TGATCAACAATTTAAAGATAAGAGAAAGAAAGAAATCAGAA 1367
|||||
Db 41 TGATCAACAATTTAAAGATAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2
RESULT 5
BM560789
LOCUS BM560789 1135 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6566093 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5550711
5', mRNA sequence.
ACCESSION BM560789
VERSION BM560789.1 GI:18805500
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1135)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLMI2263 row: e column: 16
High quality sequence stop: 423.
Location/Qualifiers
1..1135

```



```
QY 866 ATCTCTGTCAGCCATCATCAACCAAAATGATGCTAGATCTTGTCTGATTCATACAAAGTG 925
Db 494 ATCTCTGTCAGCCATCATCAACCAAAATGATGCTAGATCTTGTCTGATTCATACAAAGTG 435
QY 926 AAAGAAGGCTGAATCCCTCCAAAGACCTATTCAATTAACAATAGCCCTGCGTTGCTGTT 985
Db 434 AAAGAAGGCTGAATCCCTCCAAAGACCTATTCAATTAACAATAGCCCTGCGTTGCTGTT 375
QY 986 TTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGCGGGTTATCTTATGCTCTCA 1045
Db 374 TTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGCGGGTTATCTTATGCTCTCA 315
QY 1046 TCAATCGGCTGTTTTCCTGAGTTTCCCTGAGTTTCCCTGCTGCTGCTGCTGCTGCTGCT 1105
Db 314 TCAATCGGCTGTTTTCCTGAGTTTCCCTGAGTTTCCCTGCTGCTGCTGCTGCTGCTGCT 255
QY 1106 TGAGTGGTGATGCTTTTTTACACCTTCTCCACATTTCTCATGCAAGTCACCAACATAGTC 1165
Db 254 TGAGTGGTGATGCTTTTTTACACCTTCTCCACATTTCTCATGCAAGTCACCAACATAGTC 195
QY 1166 ATAGCCATGAAGAACCAAGCAATGGAATGAAAGAGGACCACTTTTTCAGTCTATCTGCTT 1225
Db 194 ATAGCCATGAAGAACCAAGCAATGGAATGAAAGAGGACCACTTTTTCAGTCTATCTGCTT 135
QY 1226 CTCAAAACATAGAAGAAAGTGCTATTGTTGATTTCCACGTGGAAGGCTTAACAGCTCTAG 1285
Db 134 CTCAAAACATAGAAGAAAGTGCTATTGTTGATTTCCACGTGGAAGGCTTAACAGCTCTAG 75
QY 1286 GAGGCTGCTATTTTCATGTTTCTGTTGACATGCTGCTCAGTCAATGATCAACAAATTTAAAG 1345
Db 74 GAGGCTGCTATTTTCATGTTTCTGTTGACATGCTGCTCAGTCAATGATCAACAAATTTAAAG 15
QY 1346 ATAAAGAAAGAAAA 1358
Db 14 AAAAAAAGAAAAA 2

RESULT 7
BG168169
LOCUS
DEFINITION 602341563P1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4449591 5',
mRNA sequence.
ACCESSION BG168169
VERSION BG168169.1 GI:12674872
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgi.nhl.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LAM10233 row: m column: 16
High quality sequence start: 14
High quality sequence stop: 659.
Location/Qualifiers
1. .957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4449591"
/tissue_lib="NIH_MGC_89"
/tissue_type="hybernephroma, cell line"
/lab_host="DH10B (phage-resistant)"

FEATURES
source
```

```
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dr' primed.
Average insert size 1.3 Kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 321 a 229 c 204 g 203 t
ORIGIN
Query Match 27.7%; Score 627.6; DB 12; Length 957;
Best Local Similarity 89.6%; Pred. No. 4.6e-130;
Matches 731; Conservative 0; Mismatches 79; Indels 6; Gaps 5;
QY 288 CGACCATCATCAGACACGAGCATCAC-TCAGACCATGAGC-GTCACCTCAGACCATGAG 345
Db 1 CGACCATCATCAGACACGAGCATCACGTGACACCATGAGCGGTGCTCAGACCATGAG 60
QY 346 C-ATCAGTCAGACACGAGCATCACCTCTGACCATGATCATCTCTCACCATAATCATGC 404
Db 61 CAATCATCAGACACGAGCATCACCTCTGACCATGATCATCTCTCACCATAATCATGC 120
QY 405 TGC-TTCTGTTAAATAAGCGAAAGAGCTCTTTTGCCAGACCATGACTCAGATAGTTTCA 463
Db 121 TGCCTTCTGTTAAATAAGCGAAAGAGCTCTTTTGCCAGACCATGACTCAGATAGTTTCA 180
QY 464 GTAAAGATCTCTAGAAACAGCGAGGAAAGAGCTCTCAGCAGCAGAACATGCCAGTGGTA 523
Db 181 GTAAAGATCTCTAGAAACAGCGAGGAAAGAGCTCTCAGCAGCAGAACATGCCAGTGGTA 240
QY 524 GAAGGAATGTCAAGGACAGTGTGTTAGTGTAGTGAAGTGAAGTCAACTGTGTACAACACTG 583
Db 241 GAAGGAATGTCAAGGACAGTGTGTTAGTGTAGTGAAGTGAAGTCAACTGTGTACAACACTG 300
QY 584 TCTCTGAAGGAACCTCACTTCTTAGAGACATAGAGACTCCAAGACCTGGAAACACTCTTCC 643
Db 301 TCTCTGAAGGAACCTCACTTCTTAGAGACATAGAGACTCCAAGACCTGGAAACACTCTTCC 360
QY 644 CCAAGATGTAAAGCAGCTCCACCTCCAGCTGTACATCAAGAGAGCGGTGAGCGCGG 703
Db 361 CCAAGATGTAAAGCAGCTCCACCTCCAGCTGTACATCAAGAGAGCGGTGAGCGCGG 420
QY 704 TGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCGCCCGGAAAGGCTTTATGTATTCCA 763
Db 421 TGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCGCCCGGAAAGGCTTTATGTATTCCA 480
QY 764 GAAACACAAATGAAATCTCTCAGGAGTGTTCATGATCAAGAGTACTGACATCTCATG 823
Db 481 GAAACACAAATGAAATCTCTCAGGAGTGTTCATGATCAAGAGTACTGACATCTCATG 540
QY 824 GCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCCATCA 883
Db 541 GCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTGTGCCAGCCATCA 600
QY 884 TCAACCAAAATTTGATGCTAGATCTTGTGATTCATCAAGTGAAGAGAGGCTGAAATCC 943
Db 601 TCAACCAAAATTTGATGCTAGATCTTGTGATTCATCAAGTGAAGAGAGGCTGAAATCC 660
QY 944 CTCCAAAGACCTATTTCATTACAATAGCCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGT 1003
Db 661 TTCCAAAGAGCTAATTCATTACAAATAGGCTGAGTGGGTGGTGGTGGTGGTGGTGGTGGT 720
QY 1004 TCAGTTCCTGCTCTGCTGCGGGTTATCTTACAAATAGGCTGAGTGGGTGGTGGTGGTGGT 1063
Db 721 ATCAGTACCTG--CACTGGCGGGGTAAACTAAACGGCCCTTAATGATCGGGGTGCCACA 778
QY 1064 AATTCCTCCTGAGTGTTCCTTGTGGCAGCTGGCCGCTTG 1099
Db 779 AATCCCTGGCATTCTCTGTGACTGCGCTGGCAGCTTG 814

RESULT 8
BG530601
LOCUS
DEFINITION 602560191F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:469787 5',
```

Accession	BG530601	GI:13522138	mRNA sequence.
Version	BG530601.1		
Keywords	EST.		
Source	human.		
Organism	Homo sapiens		
Reference	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
Authors	J (bases 1 to 746)		
Title	NIH-MGC http://mqc.nci.nih.gov/.		
Journal	National Institutes of Health, Mammalian Gene Collection (MGC)		
Comment	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: rquapbs@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: CLONETECH Laboratories, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLCMI527 row: f column: 23		
	High quality sequence stop: 623.		
Features	Location/Qualifiers		
Source	1..746		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4697878"		
	/clone_lib="NIH_MGC_61"		
	/tissue_type="embryonal carcinoma"		
	/lab_host="DIII0B (Ti phage-resistant)"		
	/note="Organ: testis; Vector: pDNR-LIH (Clontech); Site_1: SfiI (ggcgctcgccg); Site_2: SfiI (ggcattatgcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCCGACATG-TT(30)NN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."		
Base Count	195 a 151 c 179 g 221 t		
Query Match	26.5%	Score 600.2; DB 12; Length 746;	
Host local Similarity	96.4%	Pred. No. 6.1e-124;	
Matches	647; Conservative	0; Mismatches 18; Indels 6; Gaps 3;	
QY	1575	GGAAGTCTACAATCAATATGTATCCACAGGGTGCAGAAATAAATGCCATTTCACATTTC	1634
DB	1	GGAAGTCTACANTGAATATGTATCCACAGGGTGCAGAAATAAATGCCATTTCACATTTC	60
QY	1635	CGATPACACTCGGCCAGTCAGACGATCTCATTCACCAACCATCATGACTACCAATCATTTCT	1694
DB	61	CGATACACTCGGCCAGTCAGACGATCTCATTCACCAACCATCATGACTACCAATCATTTCT	120
QY	1695	CCATCATCACCACC-ACCAAAACCAACCATCTCACAGTCACAGCGAGCTACTCTCGG	1753
DB	121	CCATCATCACCACCAGCAAAACCAACCATCTCACAGTCACAGCGAGCTACTCTCGG	180
QY	1754	AGAGCTCGAAAGATGCCGGCGTCGCCACTTTGGCGCTGGATGATGATGATGATGATGATG	1813
DB	181	AGGAGCTCGAAAGATGCCGGCGTCGCCACTTTGGCGCTGGATGATGATGATGATGATG	240
QY	1814	TGCACAAATTTTACGCGATGGCCTAGCAATGGTGTGCTTTTACTGAAGGCTTATCAAGT	1873
DB	241	TGCACAAATTTTACGCGATGGCCTAGCAATGGTGTGCTTTTACTGAAGGCTTATCAAGT	300
QY	1874	GTTTTAAGTACTTCTGTTGCTGTGTTCTGTCATGAGTTGCGCTCATGAATTAAGTACATTTG	1933
DB	301	GTTTTAAGTACTTCTGTTGCTGTGTTCTGTCATGAGTTGCGCTCATGAATTAAGTACATTTG	360

pluescript SK-⁺ by Dr. H. Inoue following the Washington University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

BASE COUNT 163 a 143 c 130 g 168 t

Query Match 26.3%; Score 597.6; DB 14; Length 604;
Best Local Similarity 99.3%; Pred. No. 2.3e-123;
Matches 600; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	537	GCACAGCTGTATGCTAGTGAAGTCACTCAACTGTGTACACACTGTCTCTGAAGAAC	596
Db	1	GCACAGGTTTACTGCTAGTGAAGTCACTCAACTGTGTACACACTGTCTCTGAAGAAC	60
QY	597	TCACCTTTCTAGACAAATAGACACTCCAAAGACCTCGAAACCTCTTCCGCCAAGATGTAA	656
Db	61	TCACCTTTCTAGACAAATAGACACTCCAAAGACCTCGAAACCTCTTCCGCCAAGATGTAA	120
QY	657	CAGCTCCACTCCACCCAGCTGTACATCAAGAGCCGGGTGAGCCGGTGGCTGTAGGAA	716
Db	121	CAGCTCCACTCCACCCAGCTGTACATCAAGAGCCGGGTGAGCCGGTGGCTGTAGGAA	180
QY	717	ACAATGAATCTGTGAGTGAGCCCGGAAAGGCTTTATGTATTCGAAACACAATGA	776
Db	181	ACAATGAATCTGTGAGTGAGCCCGGAAAGGCTTTATGTATTCGAAACACAATGA	240
QY	777	AAATCTCAGGAGTCTTCAATGCATCAAGCTACTGACATCTCATGCGATGGGATCA	836
Db	241	AAATCTCAGGAGTCTTCAATGCATCAAGCTACTGACATCTCATGCGATGGGATCA	300
QY	837	GGTTCGGCTCAATGCAACAGAGTTCACATCTCTGTCCAGCCATCATCAACAAATGA	896
Db	301	GGTTCGGCTCAATGCAACAGAGTTCACATCTCTGTCCAGCCATCATCAACAAATGA	360
QY	897	TGCTAGATCTCTGATCATCAAGTCAAGAAAGAGGCTGAAATCCCTCCAAACACTA	956
Db	361	TGCTAGATCTCTGATCATCAAGTCAAGAAAGAGGCTGAAATCCCTCCAAACACTA	420
QY	957	TTTCATTTACAAATAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	1016
Db	421	TTTCATTTACAAATAGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT	480
QY	1017	TCGCTGGGGGTATCTTAGTGCCTCTCATGAATCGGCTGTTTTCACAAATTTCTAG	1076
Db	481	TCGCTGGGGGTATCTTAGTGCCTCTCATGAATCGGCTGTTTTCACAAATTTCTAG	540
QY	1077	TTTCTCTGCGCACTGGCGGTTGGGACCTTTGAGTGGTGGTGGTGGTGGTGGTGGT	1136
Db	541	TTTCTCTGCGCACTGGCGGTTGGGACCTTTGAGTGGTGGTGGTGGTGGTGGTGGT	600
QY	1137	ACAT 1140	
Db	601	ACAT 604	

RESULTS 10
BM948026 766 bp mRNA linear EST 14-MAR-2002
LOCUS UI-M-EG0P-bvd-b-19-0-01.r1 NIH_BMAP_EG0P Mus musculus cDNA clone
DEFINITION IMAGE:5690418 5', mRNA sequence.
ACCESSION BM948026
VERSION BM948026.1 GI:19431616
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 766)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP).
The following repetitive elements were found in this cDNA sequence: 318-399, >(TGG)n#Simple_repeat (matched complement)
Seq primer: pYX-5.

FEATURES
source 1..766
Location/Qualifiers
/organism="Mus musculus"
/strain="c57BL/6"
/db_xref="taxon:10090"
/clone_image="5690418"
/clone_lib="NIH_BMAP_EG0P"
/tissue_type="whole brain"
/dev_stage="embryo.18.5 dpc"
/lab_host="DH10B (11 phage resistant)"
/note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I; Site_2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCAGCAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 201 a 208 c 192 g 165 t

Query Match 25.9%; Score 586.8; DB 14; Length 766;
Best Local Similarity 85.4%; Pred. No. 6.2e-121;
Matches 654; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1348	AAGAAGAAAAGAAATCAGAAGAACCTTGAATCATCATGTGGAGATTAAAGACGAG	1407
Db	1	AAGAAGAAAAGAAATCAGAAGAACCTTGAATCATCATGTGGAGATTAAAGACGAG	60
QY	1408	TTGTCCAAGTATGAATCTCAACTTTTCAACAATGAGGAGAAAGTAGATACAGATGAT	1467
Db	61	CTGTCCAATACGACTCTCAGCTTTCTCTCAATGAGAGAAAGGTGGACCCAGGGAAC	120
QY	1468	ACTGAAGGCTATTACGAGCAGACTCACAAGAGCCCTCCACCTTTGATTTCTCAGCAG	1527
Db	121	CCTGAAAGCTATCTGCGAGCGGACATCCCAAGAGCCCTTCCCTTTGATTTCTCAGCAG	180
QY	1528	GCAGTCTTGAAGAAAGAGGTCTCATAGCTCATCATCCACAGGAGGTCTACAAT	1587
Db	181	ACGATGTTGAGAGGAGAGGTCTCATAGCTCATCATCCACAGGAGGTCTACAAT	240
QY	1588	GAATATGTACCCAGAGGGTGCAGAATAAATGCCATTTACATTTTCCACGATACATCGG	1647
Db	241	GAATATGTGCCCAGGGCTGCAAGAACAAAGTGCCATTCACATTTCCACGATACGCTGG	300
QY	1648	CAGTCACAGCATCTCATTTCCACCACCATCATGACTACCATCATATTTCTCCATCATC	1707
Db	301	CAGTCACAGCATCTCATTTCCACCACCATCATGACTACCATCATATTTCTCCATCATC	360

```
Qy 1708 CACCAAAACCACTCTCACAGTCCACAGCCAGCGCTACTCTGGGAGGAGK: TGAAGAT 1767
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CACCAGAACCAACCCCTCACAGCCACAGCCAGCGCTACTCTCGAGAGGAGTGAAGGAC 420

Qy 1768 GCGGGCTGCCACTTTGGCTGGATGGTGATTAATGGGTGATGGCCCTGCACAAATTTCAAC 1827
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCGGGATTTCCACATTTGGCTGGATGGTGATTAATGGGTGATGGCCCTGCACAAATTTCAAC 480

Qy 1828 GATGGCTACCAATTTGCTGCTCTTTACTGAAGGCTTATCAAGTGGTTTAAGTACTTCT 1887
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 GACGGCTTCTCTATTTGGTGGCTCTTCAACGAGGTTTGTCCAGTGGCTTAAGCAGCTCT 540

Qy 1888 GTTTCCTGTTCTCTCATGAGTTTCCCTCATGAATTAAGTCACTTTGCTCTCTTCACTAA 1947
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GTCGCTGTTCTCTCATGAAGTGCCTCATGAAGTCACTAGTGCATTTGCTGTTTGGCTAAAG 600

Qy 1948 GCTGGCATGACGTTTAAGCAGGCTGCTCTTTATTAATGCAATGTCAGCCATCTGGCGTAT 2007
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 GCTGGCATGACTCTCAAGCAGGCTGCTCTTAATGCTCTGTCAGCCATTTGGCGCTAC 660

Qy 2008 CTTTGAATGCAACAGCAATTTTTCATTGGTCAATTATGCTGCAAAATGTTTCTATCTGATA 2067
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CTTTGAATGCAACAGGATATTATCATCGGCGATTAATGCAAGAAATGCTGCTATGCTGATA 720

Qy 2068 TTTTGCATTTCTGCTGCTTATTCATGATGTTGCTCTGCTGCTGATA 2113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 TTTGCACTGCTGCGGCTTTGTTTCATGATGCTGCTGCTGCTGCTGATA 766

RESULT 11
HF983458 925 bp mRNA linear EST 23-JAN-2001
602307085F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398429 5',
mRNA sequence.
HF983458
HF983458.1 GI:12386270
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 925)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-f@mail.nih.gov
Tissue procurement: ATCC
cDNA library preparation: Life Technologies, Inc.
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10100 row: 1 column: 22
High quality sequence stop: 631.
Location/Qualifiers
1..925
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4398429"
/tissue_type="NIH_MGC_88"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: small intestine; Vector: pCMV-Sport6;
Site:1: Noli; Site:2: Sali; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 223 a 222 c 215 g 265 t
ORIGIN
Query Match 25.4% Score 576; DB 12; Length 925;
```

```
Best Local Similarity 94.8%; Pred. No. 1.7e-118;
Matches 660; Conservative 0; Mismatches 30; Indels 6; Gaps 6;

Qy 1293 GTATTTCATGTTCTTGTGAACATGCTCTCACATTTGATCAACAATTTAAGATAAGAA 1352
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GTATTTCATGTTCTTGTGAACATGCTCTCACATTTGATCAACAATTTAAGATAAGAA 60

Qy 1353 GAAAAGAAATCAGAAAGAAACCTTGAATATGATGATGATGATGATGATGATGATGATGAT 1412
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 GAAAAGAAATCAGAAAGAAACCTTGAATATGATGATGATGATGATGATGATGATGATGAT 120

Qy 1413 CAAGTATGAATCTCAACTTTCAACAAATGAGGAGAAATAGATACAGATGATGATGATGAT 1472
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CAAGTATGAATCTCAACTTTCAACAAATGAGGAGAAATAGATACAGATGATGATGATGAT 180

Qy 1473 AGGCTATTACAGCAGAGACTCACAGAGCCCTCCACATTTGATTTCTCAGCAGCTGCACT 1532
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 AGGCTATTACAGCAGAGACTCACAGAGCCCTCCACATTTGATTTCTCAGCAGCTGCACT 240

Qy 1533 CTTTGAAGAAGAAAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATA 1592
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CTTTGAAGAAGAAAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAATGAATA 300

Qy 1593 TGTACCCAGAGGGTGCAGAATAAATGCCATTCACATTTCCACAGATACACTCGGCCAGTC 1652
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TGTACCCAGAGGGTGCAGAATAAATGCCATTCACATTTCCACAGATACACTCGGCCAGTC 360

Qy 1653 ACACGATCTCATTCACACCATCATGACTACCATCATATCTCCATCATCACACACCA 1712
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 AGACGATCTCATTCACACCATCATGACTACCATCATATCTCCATCATCACACACCA 420

Qy 1713 AAACACCATCTCATGACTCACAGCCAGCG-CTACTCTCGGAGGAGCTGAAAGATGCCG 1771
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 AAACACCATCTCATGACTCACAGCCAGCGACTACTCTCGGAGGAGCTGAAAGATGCCG 480

Qy 1772 GCGTCGCCACTTTGGCTGGATGATTAATGGGTGATGGCTGCACAA-TTTCAGCAT 1830
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 GCGTCGCCACTCTGGCTGGATGATTAATGGGTGATGGCTGCACAA-TTTCAGCAT 540

Qy 1831 GCGCTAGCAATTTGG-TGCTGCTTTTACTGAA-GGCTTATCAAGTGGTTTAAGTACTTCTG 1888
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 GCGCTAGCAATTTGGCTGGATGATTAATGGGTGATGGCTGCACAA-TTTCAGCAT 600

Qy 1889 TTGCTGTGTTCTGTCATGAGTTTCCCTCATGAATTAAGTGAATTTGCTTACTTAAAG 1948
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 -TGCTGTGTTCTGTCATGAGTGCCTCATGAATTAAGTGAATTTGCTTACTTAAAG 659

Qy 1949 CTGGCATGACCGTTAAGCAGGCTGCTTTTATAATG 1984
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 660 GTGGCATTACCG-TAAGCGCGCTGCTTTCATG 694

RESULT 12
AI907176 743 bp mRNA linear EST 30-MAR-2000
LOCUS RC-BT133-190399-076 BT133 Homo sapiens cDNA, mRNA sequence.
DEFINITION AI907176
ACCESSION AI907176
VERSION AI907176.1 GI:6497502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 743)
DIAS Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL
```


[illegible]

|||||
Db 121 TCCTGCTCTCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 180
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1129
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 240
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1189
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 300
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1249
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 360
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1309
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 420
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1369
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 480
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1429
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 540
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1489
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 600
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1548
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 660
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1605
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 720
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1665
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 779
QY TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 1707
Db TCCTGAGTTCTCTGCTGCTGCTGGAGTCAATCTTGGTGGCACTCATCAACCGGGTATTTTCAAGTTCC 821

RESULT: 14
BM963737
LOCUS
DEFINITION
UI-M-EQ0-bwn-g-10-0-UI.r1 NIH_BMAP_EQ0 Mus musculus linear EST 18-MAR-2002
IMAGE:5698593 5', mRNA sequence.
ACCESSION
BM963737
VERSION
BM963737.1 GI:19547157
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 751)
REFERENCE
1 (bases 1 to 751)
AUTHORS
NTH-MGC <http://mhc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: craps-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
The following repetitive elements were found in this cDNA
sequence: 415-496, >(TGG)nSimple_repeat (matched complement)
Seq primer: pX-5.
FEATURES
source
1. .751
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5698593"
/clone_lib="NIH_BMAP_EQ0"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="D10108 (T1 phase resistant)"
/note="Organ: brain; Vector: pX-Asc; Site: 1; EcoR I; Site: 2; Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCAGCAG. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."
BASE COUNT 200 a 203 c 183 g 163 t 2 others
Query Match 25.0%; Score 566.8; DB 14; Length 751;
Best Local Similarity 85.5%; Pred. No. 1.9e-116;
Matches 642; Conservative 0; Mismatches 108; Indels 1; Gaps 1;
QY 1252 TTTGATTCCACGT-GGAAGGCTCTAACAGCTCTAGAGGCGCTGATTTCATGTTCTTGT 1310
Db 1 TTTGATTCCACGTNGGAAAGGCTGACGCTCTAGGGGCTTATATTTCATGTTCTTGT 60
QY 1311 TGAACATGCTCTCACATTGATCAACAATTTAAAGATAGAAGAAAAGAAATCAGAAGA 1370
Db 61 GGAACACGCTACTCACACTGATCAAGCAATTTAAAGATAAGAAAAGAAATCAAGAAG 120
QY 1371 ACCTGAAATGATGATGTGGAGATTAAGACAGTGTGCCAAGTATGAATCTCAACT 1430
Db 121 ACCTGAAATGATGAGGATGTGGAGACAGACAGCTGTCCAAATACCACTCTCAGCT 180
QY 1431 TTCACAAATGAGAGAAAGTAGATACAGATGATCGAAGTGAAGGCTATTTACGAGCAGA 1490
Db 181 TTCCTCAATGAGAGAAAGGTGACCCAGGGAAACGCTGAAAGCTATCTGGAGCCGA 240
QY 1491 CTCACAGAGCCCTCCCACTTTGATTCTCAGAGCCTGCGAGTCTTGAAGAGAAAGGT 1550
Db 241 CTCACAGAGCCCTCCCACTTTGATTCTCCAGCAGCGAGGATGTTGGAAGAGAAAGGT 300
QY 1551 CATGATAGCTCATCTCATCCACAGGAAGTCTACAATGAATATGTACACAGAGGTTGCAA 1610
Db 301 CATGATAGCCCTGACACACCCACAGAAAGTCTACAATGAATATGTCCCAAGGGCTGCAA 360
QY 1611 GAATAATGCCATTACATTTTCCACGATACACTTCGCGCCAGTCAGAGGATCTCATTCACCA 1670
Db 361 GAACAAGTGCCATTACACTTCCACGATACGCTGGCCAGCTCGAGACCTCATCCACCA 420
QY 1671 CCATGATGATACCATCATATTTCTCCATCATATCACACACCAACCAACCATCTCCACAG 1730
Db 421 CCATGATGATACCATCATATTTCTGACACCAACCAACCAACCAACCAACCAACCAAC 480
QY 1731 TCACAGCCAGCGCTACTCTCGGGAGGAGGTGAAAGATCGGGGTGCGGCTCTTGGCCTG 1790
Db 481 CCACAGCCAGCGCTACTCTCGAGAGGAGGCTGAAGAGCGCGGCAATTTGCCACATTTGGCCTG 540

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 04:15:57 ; Search time 143 seconds

(without alignments)

6107.985 Million cell updates/sec

Title: us-09-642-034-4

Perfect score: 2268

Sequence: 1 atggcgaggagattatctgt.....tggttcgtataaatlctag 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/us07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1141.4	50.3	1193	10	US-09-925-300-591 Sequence 591, App
2	319.2	14.1	337	9	US-09-920-455-250 Sequence 250, App
3	319	14.1	5330	10	US-09-789-561-14 Sequence 14, Appl
4	127.6	5.6	4573	10	US-09-880-107-1649 Sequence 1649, Ap
5	62.2	2.7	574	10	US-09-864-761-228 Sequence 228, App
6	62.2	2.7	669	10	US-09-864-761-17051 Sequence 17051, A
7	59.8	2.6	285	9	US-09-286-488-64 Sequence 64, Appl
8	59.8	2.6	285	10	US-09-737-178-64 Sequence 64, Appl
9	59.8	2.6	342	9	US-09-286-488-65 Sequence 65, Appl
10	59.8	2.6	342	10	US-09-737-178-65 Sequence 65, Appl
11	59.8	2.6	351	9	US-09-286-488-60 Sequence 60, Appl
12	59.8	2.6	351	10	US-09-737-178-60 Sequence 60, Appl
13	59.8	2.6	356	9	US-09-286-488-63 Sequence 63, Appl
14	59.8	2.6	356	10	US-09-737-178-63 Sequence 63, Appl
15	59.8	2.6	358	9	US-09-286-488-58 Sequence 58, Appl
16	59.8	2.6	358	10	US-09-737-178-58 Sequence 58, Appl
17	59.8	2.6	363	9	US-09-286-488-66 Sequence 66, Appl
18	59.8	2.6	363	10	US-09-737-178-66 Sequence 66, Appl
19	59.8	2.6	363	10	US-09-737-178-66 Sequence 66, Appl

c 20	59.8	2.6	363	10	US-09-737-178-67 Sequence 67, Appl
c 21	59.8	2.6	409	9	US-09-286-488-59 Sequence 59, Appl
c 22	59.8	2.6	409	10	US-09-737-178-59 Sequence 59, Appl
c 23	59.8	2.6	410	9	US-09-286-488-61 Sequence 61, Appl
c 24	59.8	2.6	410	10	US-09-737-178-61 Sequence 61, Appl
c 25	59.8	2.6	416	9	US-09-286-488-62 Sequence 62, Appl
c 26	59.8	2.6	416	10	US-09-737-178-62 Sequence 62, Appl
c 27	59.8	2.6	1271	9	US-09-286-488-5 Sequence 5, Appl
c 28	59.8	2.6	1271	10	US-09-737-178-5 Sequence 5, Appl
c 29	59.8	2.6	1271	10	US-09-737-178-42 Sequence 42, Appl
c 30	59.8	2.6	1271	10	US-09-737-178-42 Sequence 42, Appl
c 31	59.8	2.6	1821	9	US-09-286-488-6 Sequence 6, Appl
c 32	59.8	2.6	1821	10	US-09-737-178-6 Sequence 6, Appl
c 33	59.4	2.6	1635	10	US-09-864-761-20241 Sequence 20241, A
c 34	59.4	2.6	1973	10	US-09-864-761-3471 Sequence 3471, Ap
c 35	58.8	2.6	4223	9	US-09-286-488-7 Sequence 7, Appl
c 36	58.8	2.6	4223	10	US-09-737-178-7 Sequence 7, Appl
c 37	58.8	2.6	4223	10	US-09-737-178-7 Sequence 7, Appl
c 38	58.8	2.6	4223	10	US-09-737-178-45 Sequence 45, Appl
c 39	57.8	2.5	468	10	US-09-864-761-2321 Sequence 2321, Ap
c 40	57.2	2.5	18596	9	US-09-954-531-124 Sequence 124, App
c 41	57.2	2.5	18596	10	US-09-954-531-348 Sequence 348, App
c 42	57.2	2.5	18596	10	US-09-880-107-1590 Sequence 1590, Ap
c 43	57.2	2.5	18596	10	US-09-967-768A-119 Sequence 119, App
c 44	56.8	2.5	464	10	US-09-864-761-764 Sequence 764, App
c 45	56.8	2.5	583	10	US-09-864-761-20772 Sequence 20772, A

ALIGNMENTS

RESULT 1

US-09-925-300-591
; Sequence 591, Application us/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 591
; LENGTH: 1193
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-591

Query Match 50.3%; Score 1141.4; DB 10; Length 1193;
Best Local Similarity 99.7%; Pred. No. 3.6e-308;
Matches 1139; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY	539	ACAGTCTTACTGCTAGTGAAGTGACCTCAACTGCTGTACAACTGCTCTCTGAAGGAATCTC	598
DB	1	ACAGTCTTACTGCTAGTGAAGTGACCTCAACTGCTGTACAACTGCTCTCTGAAGGAATCTC	60
QY	599	ACTTTCTAGAGACAATAGAGACTCCAAAGACCTCGGAACTCTTCCCAAGATGTAAGCA	658
DB	61	ACTTTCTAGAGACAATAGAGACTCCAAAGACCTCGGAACTCTTCCCAAGATGTAAGCA	120
QY	659	GCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGTGGCTGTAGGAAAA	718
DB	121	GCTCCACTCCACCCAGTGTACATCAAGAGCCGGGTGAGCCGGTGGCTGTAGGAAAA	180
QY	719	CAATCAATCTGTGAGTGAAGCCCGGAAAGGCTTTATGTTATTCACAGAAACACAAATGAAA	778
DB	181	CAATCAATCTGTGAGTGAAGCCCGGAAAGGCTTTATGTTATTCACAGAAACACAAATGAAA	240

QY	779	ATCCTCTAGGAGTGTGTTTCAATGCAATCAAAAGCTACTCAGCATCTCATGCGCATGCGCATCCACG	838
nb			
nb	241	ATCCTCTAGGAGTGTGTTTCAATGCAATCAAAAGCTACTGCAATCTCATGCGCATGCGCATCCACG	300
QY	839	TTCGGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCGCATCATCAACCAAAATTCATG	898
nb			
nb	301	TTCGGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCGCATCATCAACCAAAATTCATG	360
QY	899	CTAGATCTTCTGTCGATTCATCATCAAGTGAAAAGAGGCTGAAATCCCTCCAAAGACCTATTT	958
nb			
nb	361	CTAGATCTTCTGTCGATTCATCATCAAGTGAAAAGAGGCTGAAATCCCTCCAAAGACCTATTT	420
QY	959	CATTACAAATAGGCTGGGTTTGGTGGTTTATAGCCATTTCCCATCATCAGTTTCTCTGTCCTC	1018
nb			
nb	421	CATTACAAATAGGCTGGGTTTGGTGGTTTATAGCCATTTCCCATCATCAGTTTCTCTGTCCTC	480
QY	1019	TGCTGGGGGTATCTTTAGTGCCCTCTCATGAATCGGGTGTTTTTCAAAATTTCTTCTTGATTT	1078
nb			
nb	481	TGCTGGGGGTATCTTTAGTGCCCTCTCATGAATCGGGTGTTTTTCAAAATTTCTTCTTGATTT	540
QY	1079	TCCTTTGGGCACATGCGCGTTTGGGACTTTTCAGTGGTGATGCTTTTTCACACCTTTCTTCCAC	1138
nb			
nb	541	YCCYTFGGCACATGCGCGTTTGGGACTTTTCAGTGGTGATGCTTTTTCACACCTTTCTTCCAC	600
QY	1139	ATTCATCATGCAAGTCAACCACCATAGTCATAGCCATGGAAGAACCCAGCAATGGAATGAAAA	1198
nb			
nb	601	ATTCATCATGCAAGTCAACCACCATAGTCATAGCCATGGAAGAACCCAGCAATGGAATGAAAA	660
QY	1199	GAGACCACTTTTTCAGTCATCTGTCTTCTCAAAACATAGAAGAAGTGCCCTATTTTTCGATT	1258
nb			
nb	661	GAGACCACTTTTTCAGTCATCTGTCTTCTCAAAACATAGAAGAAGTGCCCTATTTTTCGATT	720
QY	1259	CCAGCTGCGAAGGCTCTTAACAGCTCTAGGAGGCTCTGATTTTCATGTTTCTTCTTCTCAACATG	1318
nb			
nb	721	CCAGCTGCGAAGGCTCTTAACAGCTCTAGGAGGCTCTGATTTTCATGTTTCTTCTTCTCAACATG	780
QY	1319	TCTCTCATCTTGTATCAACAACTTTAAAGATTAAGAAGAAAAGAAATCAGAAGAAACCTGAAA	1378
nb			
nb	781	TCTCTCATCTTGTATCAACAACTTTAAAGATTAAGAAGAAAAGAAATCAGAAGAAACCTGAAA	840
QY	1379	ATCATCATGATGTGGAGATTTAAGAGCAGTTGTCTCAAGTATGAAATCTCAACTTTTCAACAA	1438
nb			
nb	841	ATCATCATGATGTGGAGATTTAAGAGCAGTTGTCTCAAGTATGAAATCTCAACTTTTCAACAA	900
QY	1439	ATCAGCAGAAAGTAGATACAGATGATCCAACTTGAAGGCTATTTTACGAGCAGACATCACAAAG	1498
nb			
nb	901	ATCAGCAGAAAGTAGATACAGATGATCCAACTTGAAGGCTATTTTACGAGCAGACATCACAAAG	960
QY	1499	AGCCCTGCCACTTTGATTTCTCAGCAGCCCTGCCAGTCTTGGAGAAGAAAGAGTCTCATGATAG	1558
nb			
nb	961	AGCCCTGCCACTTTGATTTCTCAGCAGCCCTGCCAGTCTTGGAGAAGAAAGAGTCTCATGATAG	1020
QY	1559	CTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAAATAAAT	1618
nb			
nb	1021	CTCATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAAATAAAT	1080
QY	1619	GCCATTACATTTTCCACGATACACTCGGCCAGTCTAGACGATCTCTCATTTTCCACCATCATG	1678
nb			
nb	1081	GCCATTACATTTTCCACGATACACTCGGCCAGTCTAGACGATCTCTCATTTCCACCATCATG	1140
QY	1679	ACT 1681	
nb			
nb	1141	ACT 1143	

RESULT 2

RESULT: 2
US-09-920-455-250
: Sequence 250, Application US/09920455
: Patent No. US20020168647A1
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Fan, Lijun

```

: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
:
: TITLE OF INVENTION: AND DIAGNOSIS OF HEAD AND NECK CANCER
:
: FILE REFERENCE: 210121.540
:
: CURRENT APPLICATION NUMBER: US/09/920,455
:
: CURRENT FILING DATE: 2001-08-01
:
: NUMBER OF SEQ ID NOS: 275
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 250
:
: LENGTH: 337
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: misc_feature
:
: LOCATION: 42..43
:
: OTHER INFORMATION: n = A,T,C or G
:
: US-09-920-455-250

```

Query Match	14.1%	Score 319.2;	DB 9;	Length 337;
Best Local Similarity	98.5%	pred. NO. 3.1e-79;		
Matches 321;	Conservative	0;	Mismatches 5;	Indels 0

Qy	1333	AACAATTTAAGATTAAGAGAAAAGAAATCAGAAAGAAACCTGAAAATGATGATGATGATG	1399
Db	12	AATTCATTTAAACATAAAGAGAAAAGAAATNNGAAGAAACCTGAAAATGATGATGATG	71
Qy	1393	GAGATTTAAGAAGCAGTTGTCCAAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAAGTA	1452
Db	72	GAGATTAAAGAGCAGTTGTCCAAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAAGTA	131
Qy	1453	GATACAGATGATCGAACTCAAGGCTATTTACCAGCAGACTCACAGAGCCCTCCCACTTT	1512
Db	132	GATACAGATGATCGAACTCAAGGCTATTTACCAGCAGACTCACAGAGCCCTCCCACTTT	191
Qy	1513	GATTCTCAGCAGCCTCGACTCTTTGGAAGAAAGAGGTCATGATAGTCTATGCTCATCCCA	1572
Db	192	GATTCTCAGCAGCCTCGACTCTTTGGAAGAAAGAGGTCATGATAGTCTATGCTCATCCCA	251
Qy	1573	CAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAATAAATGCCATTTCACATTTTC	1632
Db	252	CAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAATAAATGCCATTTCACATTTTC	311
Qy	1633	CAGGATACACTCGGCCAGTCAGACGA	1658
Db	312	CAGGATACACTCGGCCAGTCAGACGA	337

```

RESULT 3
US-09-789-561-14
; Sequence 14, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043pl
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 5330
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-789-561-14

```

Query Match 14.1%; Score 319; DB 10; Length 5330;
Best Local Similarity 56.3%; Pred. No. 2.2e-78;
Matches 736; Conservative 0; Mismatches 535; Indels 36

RESULT 2
US-09-920-455-250 ; Sequence 250, A
; Patent No. US200
; GENERAL INFORMATION
; APPLICANT: Wand
; APPLICANT: Far

QY 969 AGCCCTGGGTTGGTGGTTTATATAGCAATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGT 1028
Db 1372 AGCCCTGGGTTGGTGGTTTATATAGCAATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGT 1431
QY 1029 TATCTTACTGCTCTCATGAATCGGGTCTTTTTCAAATTTCTCTGAGTTTCTTGTGGC 1088
Db 1432 GATCTTGGTCTCTATCATTAACCAAGGATGCTTCAAAATTCCTCTTACATTCCTTGTGTC 1491
QY 1089 ACTGCTCCCTTGGGACTTTAGCTGCTGATGCTTTTATACACCTTCTTCCACATCTCTATGC 1148
Db 1492 ATTAGCTCTAGCAACATGAGTGGAGCGCTTCTTCTCATCTAGTCCGCCATCTCTCAGG 1551
QY 1149 AAGTCACACATATAGTCATATGCCATGAAGAACACCAATGGAATGAAAGAGAGGACCTAC 1208
Db 1552 TGGACATGATCACAGTCACCAACATGCACA-----TGGGCAATGCACATTC 1596
QY 1209 TTTTCAGTCATCTGCTCTTCTCAAAACATAGAGAAAGTGCCTATTTTGAATTCACGCTGGAA 1268
Db 1597 TCATGCAATGAATCTAACAAGTTTTCGGAAGAA-----TATCATGCTCTATTTCAA 1647
QY 1269 GGTCTTAACAGCTTAGCAGGCTGTATTTTCATCTTTCTTGAACATGCTCTCACATT 1328
Db 1648 AGGACTTCTGCTCTAGAGGCACTTACTTGTCTATTTATCATTGAACACTGCGATTAAGAT 1707
QY 1429 GATCAAAACATTTAAAGATAAGA---AGAAAAGAAATCAGAGAAACCTGAAATATATGA 1385
Db 1708 CTTTAAGCACTACAAACAAACAGAGAAACAGAAATGCTTTTATGAAATAGAAACAGAA 1767
QY 1486 TCATCTGAGAGATTAAGAGCACTTCTCCAAGTATGAATCTCAACTTTTCAACAATCAGGA 1445
Db 1768 AGAATCAACTTATTGGAGAAAGCTTTCAGATCAACAAGTTAAACAAATACACAGAAATCTGA 1827
QY 1446 CAAAGTAGATACATGATCAACCTGAAGGCTATTTACGAGCAGCTTCAAGAGAGCTTC 1505
Db 1828 CTGGCTTTCACCTCAAGCTCTTGGCGAAGTATGATCACTCGGTGTTCTTGAAGATCAACT 1887
QY 1506 CCACCTTTGATCTCAGGAGCGCTGCAGTCTTGGAGAGAGAGAGGTGATGATAGTCTATGC 1565
Db 1888 TAATGAAATTCGAATGACATTTAGAAGGCCAACAAGATTCCTTAAATAATTAAGCT 1947
QY 1566 TCATCTCACAGCAAGCTTACATGAATATGTACCCAGAGGTGCAAGAAATTAATTCATATTC 1625
Db 1948 TTGTATAGAGAGAGAG-AAAATCATAGACCATTTCTCACAGTGAATGAGATTAATACATTC 2005
QY 1626 ACATTTTCACGATACACTCGGCCAGTCAGACGATCTCATTTCACTCACCATCATGACTAGCA 1685
Db 2006 -CATGAGCATGATTTCCATGCTGCTGCACATACCCAGCCAGGCAACAAAATCTGCT 2064
QY 1686 TCATATTTCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1745
Db 2065 GAGCAAGCATAAATCACCAAGTGGCACCACAAGCATTTCTCATCTCCCATGGCCCTCTGCA 2124
QY 1746 CTCTCTGGAGAGGCTGAAGATGCGCGCTGCGGCACTTTGGCCCTGGATGGTATATATGCG 1805
Db 2125 TTCTGTGATCCGCAATGGAAGAAACAGAAATAGCTAATATAGCCTGGATGCTCATCATGG 2184
QY 1806 TGATGGCTCTGCACAAATTTACAGCATGCGCTAGCAATTTGGTGGCTGCTTTTAACTGAAAGCT 1865
Db 2185 GGATGGATCCACAACCTTCACTGATGGCTGCGCAATTTGGTGGAGCTTTTCACTGCTGAT 2244
QY 1866 ATCAAGTCTGTTTAAATGATCTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1925
Db 2245 GACAGGAGGAATCAGTACTTCTATAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2304
QY 1926 TGACCTTTGCTGCTTCTTAAAGGCTGGCATGACGCTTAAAGCAGGCTGCTGCTGCTGCTGCT 1985
Db 2405 AGATTTTGGCAATTTCTTAAAGCAGGCAATGACTGTAAAGCAAGCAATTTCTATATCAACT 2464
QY 1986 ATTTCTCAGCCATGCTGGCGTATCTTGAATGGCAACAGGAATTTTCACTGCTGCTGCTGCT 2045
Db 2465 CCTCTCTGCCCATGAGCTTACATAGGCAATGCTCATAGGCAACAGCTGCTGCTGCTGCTGCT 2424

QY 2046 TGAATAATGTTCTTCTATGTTGATATTTGCACATTTACTGCTGGCTATTTTCATGTTGCTCT 2105
Db 2425 CAATAACATCACACTTTTGGATCTTTGCAGTCACTGCGAGCATGCTTCTCTATGTAACCTT 2484
QY 2106 GGTGATATGTTACCTTGAATGCTGCACAAATGATG-----CTAGTGACCATCGATGTAG 2159
Db 2485 GGTGATATGCTTCCAGAAATGTTGCATGCTGATGTCACAATGAAGAACAATGGCTTTTG 2544
QY 2160 CCGCTGGGGTATTTCTTTTACAGAAATGCTGGATGCTTTGGGTTTGGGAATATGTT 2219
Db 2545 TCTCTGGGGCAATTCATCTCTCAGAAATTTAGGATTTGCTCTTTGGATTTGGCAATATGCT 2604
QY 2220 ACTTATTTTCCATATTTGAACATAAAATCGTGTGTTTCGTATATAAAATTTCT 2266
Db 2605 GGTGATTTGCCCTCTATGAGATAAAATTCGTGTTGACATCCACCTTTT 2651
RESULT 4
US-09-880-107-1649
; Sequence 1649, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 4921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1649
; LENGTH: 4573
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 D31887
US-09-880-107-1649

Query Match 5.6%; Score 127.6; DB 10; Length 4573;
Best Local Similarity 59.6%; Pred. No. 4,7e-25;
Matches 236; Conservative 0; Mismatches 154; Indels 6; Gaps 1;
QY 1732 CACAGCCAGCGCTACTCTCGGGAGGAGCTGAAAGATGCGCGCTGCCACTTTGGCCTGG 1791
Db 1074 CAGACTGCTTGTCTACTGGCTGAAAGGTGTCGCCCTACTCTGATATCGGCACTCTGGCCTGG 1133
QY 1792 ATGGTGATTAATGGGTATGCGCTGCACAAATTTACGCGATGGCTAGCAATTTGGTCTGCT 1851
Db 1134 ATGATCAGCTCTGAGGACGCGCTCCACAATTTTCATCGATGGCTGCCACTGCTGCTTCC 1193
QY 1852 TTTACTGAAGGCTTATCAAGTGGTTTAACTACTTCTGTTGCTGTTCTGCTCATGCTTG 1911
Db 1194 TTCACTGTCTAGTTTTCCAAGGCAATCAGCACTCGGTGGCCATCTCTGTGAGGAGTTC 1253
QY 1912 CCTCATGAATTAGTGACTTTTCTGCTTACTTAAAGGCTGGCATGACCTTAAAGCAGCT 1971
Db 1254 CCACATGAGCTAGAGAGACTTTGTCATCTGCTCAAGCTGGGATGAGCATCCACAAGCT 1313
QY 1972 GTCTTTTAAATGCAATGTGACGCCATGCTGCGGTATCTTTGGAATGGCAACAGGAATTTTC 2031
Db 1314 CTCTCTTCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1373
QY 2032 ATTTGTCATATGCTCAAAATGTTTCTATGTCGATATTTTGGCACTTTACTGCTGGCTATTTC 2091
Db 1374 GCGGG-----CAGGCACCTTCTCTGCCAACGGATTTTGGCGCTAGCTGGAGGAATGTT 1427
QY 2092 ATGTATGTTGCTCTGTTGATATGCTACCTGAAATG 2127

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:23:30 ; Search time 439 Seconds
(without alignments)
11634.470 Million cell updates/sec

Title: US-09-642-034-4

Perfect score: 2268
Sequence: 1 atggcgaggaggttatctgt.....tggttcgtataaattttctag 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2268	100.0	3461	24	Prostate cancer-as
2	2268	100.0	3461	24	Human breast cancer
3	2264.8	99.9	2776	22	Human LIV-1-164647
4	2220.4	97.9	3523	23	Human prostate exp
5	2198	96.9	3461	22	Human LIV-1 gene.
6	1263	55.7	1310	19	Partial sequence o
7	1263	55.7	2404	14	Partial sequence o
8	1259.8	55.5	1310	17	Oestrogen-regulate
9	1259.8	55.5	2404	17	Oestrogen-regulate

10	1141.4	50.3	1193	21	AAF16156	Human prostate can
11	787.4	34.7	20778	22	AAK79819	Human immune/haema
12	619.6	27.3	899	24	ABA92300	Mouse breast cancer
13	549.4	24.2	551	22	AAD13487	Human LIV-1-164647
14	509.4	22.5	1597	20	AAZ33622	Human breast tumou
15	390.4	17.2	424	22	AAAL19686	Human breast cancer
16	380.4	16.8	419	22	AAAL22452	Human breast cancer
17	319.2	14.1	337	24	ABK54054	Human head and nec
18	319	14.1	3309	23	ABV23278	Human prostate exp
19	319	14.1	3309	23	ABV29124	Human prostate exp
20	319	14.1	5322	24	ABN59760	Novel human coding
21	319	14.1	5330	22	AAF91861	Human secreted pro
22	262.4	11.6	423	22	AAAL23559	Human breast cancer
23	251.4	11.1	469	22	AAAL10106	Human breast cancer
24	247	10.9	736	22	AAK92049	Human cDNA 5'-end
25	247	10.9	736	22	AAK93710	Human cDNA clone r
26	247	10.9	2863	22	AAK94381	Human full-length
27	240.8	10.6	413	22	AAAL19430	Human breast cancer
28	240.8	10.6	455	22	AAAL09043	Human breast cancer
29	224.8	9.9	406	23	ABV60690	Human prostate exp
30	206.6	9.1	643	22	AAAL18330	Human breast cancer
31	189.4	8.4	405	22	AAAL18310	Human breast cancer
32	157.4	6.9	233	23	ABV38837	Human prostate exp
33	157	6.9	176	22	AAAL17684	Human breast cancer
34	147.2	6.5	567	23	ABV08950	Human prostate exp
35	146.4	6.5	2055	23	ABL11749	Drosophila melanog
36	146	6.4	193	22	AAAL10609	Human breast cancer
37	144	6.3	4370	23	ABL11748	Drosophila melanog
38	143.8	6.3	183	22	AAAL16763	Human breast cancer
39	141.4	6.2	409	22	AAAL08417	Human breast cancer
40	140.4	6.2	207	22	AAAL18232	Human breast cancer
41	140.4	6.2	378	22	AAAL26162	Human breast cancer
42	128.4	5.7	554	22	AAAL13583	Human breast cancer
43	127.6	5.6	4573	20	AAAX33946	Human HCMV Inducib
44	127.6	5.6	4573	24	ABN95151	Gene #1649 used to
45	127.6	5.6	4631	24	ABK63926	cDNA encoding huma

ALIGNMENTS

RESULT 1

ABK92205	ID	ABK92205	standard; DNA; 3461 BP.
XX	AC	ABK92205;	
XX	DT	15-AUG-2002	(first entry)
XX	DE	Prostate cancer-associated DNA sequence #91.	
XX	DE	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;	
KW	KW	gene therapy; gene; ds.	
XX	OS	Mammalia.	
XX	PN	WO200230268-A2.	
XX	PD	18-APR-2002.	
XX	PF	12-OCT-2001; 2001WO-US32045.	
XX	PR	13-OCT-2000; 2000US-0687576.	
PR	PR	08-DEC-2000; 2000US-0733288.	
PR	PR	08-DEC-2000; 2000US-0733742.	
PR	PR	24-JAN-2001; 2001US-263957P.	
PR	PR	16-MAR-2001; 2001US-276791P.	
PR	PR	16-MAR-2001; 2001US-276888P.	
PR	PR	06-APR-2001; 2001US-281922P.	
PR	PR	24-APR-2001; 2001US-286214P.	
PR	PR	30-APR-2001; 2001US-0847046.	
PR	PR	04-MAY-2001; 2001US-288589P.	

PA (EUSB-) EOS BIOTECHNOLOGY INC.
XX Gish KC, Meck DH, Wilson KB, Afar D, Hevezi P;
XX WPI: 2002-471335/50.
XX P-PSDB; ABG61889.
XX Detecting a prostate cancer-associated transcript in a cell in a
XX patient, useful for diagnosing prostate cancer (PC) or screening
XX modulators of PC, by determining if prostate cancer-associated genes
XX are expressed in a prostate tissue
XX Claim 22: Page 374-375; 436pp; English.
XX
XX The present invention relates to methods of detecting a prostate
XX cancer-associated transcript in a cell from a patient. The method
XX comprises contacting a biological sample from the patient with
XX prostate cancer-associated polynucleotides (designated PC genes) that
XX selectively hybridize to a sequence that is at least 80% identical
XX to them. The prostate cancer-associated polynucleotide sequences
XX are differentially expressed in prostate tumour tissue or in
XX prostate cancer and are derived from the tissues of various
XX organisms such as humans or other mammals (e.g. mice, sheep and dogs).
XX The methods of the invention are useful for diagnosing and treating
XX prostate cancer in mammals. The prostate cancer-associated genes are
XX useful for diagnosing or treating prostate cancer, as well as for
XX identifying modulators of prostate cancer or agents that inhibit
XX prostate therapy. The nucleic acid sequences are particularly useful
XX in gene therapy, as a vaccine or in antisense applications.
XX AHK92115-AHK92263 represent prostate cancer-associated polynucleotide
XX sequences.
XX
XX Sequence 3461 BP: 1042 A: 686 C: 728 G: 1005 T: 0 other:
XX
XX Query Match 100.0%; Score 2268; DB 24; Length 3461;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ATGCGGAGGAGTTATCTGTATCTTGATCCTGACCTTGGCCCTCTCTCTCAAAATCCC 60
DB 138 ATGCGGAGGAGTTATCTGTATCTTGATCCTGACCTTGGCCCTCTCTCTCAAAATCCC 197
QY 61 CTTCATGAATAAAGCAGCTGCTTTCCGCCAGACCTGAGAAATAGTCGGAATCG 120
DB 198 CTTCATGAATAAAGCAGCTGCTTTCCGCCAGACCTGAGAAATAGTCGGAATCG 257
QY 121 GAATCTGGCATTAATCTTGACCTTGGCAATTTCCACAGCGCAATATCATCTACAAACAGCTT 180
DB 258 GAATCTGGCATTAATCTTGACCTTGGCAATTTCCACAGCGCAATATCATCTACAAACAGCTT 317
QY 181 TTCTACCGCTATGGAGAAATTAATCTTTGTGTCAGTTGAAGGTTTCAGAAAATTAATCTTCAA 240
DB 318 TTCTACCGCTATGGAGAAATTAATCTTTGTGTCAGTTGAAGGTTTCAGAAAATTAATCTTCAA 377
QY 241 AATATAGCATAGATAAGATTAAAGAAATCCATATACACCATGACCAGCAGCATCAGTCA 300
DB 378 AATATAGCATAGATAAGATTAAAGAAATCCATATACACCATGACCAGCAGCATCAGTCA 437
QY 301 GACCAGGAGCATCAGTCAAGACCATGAGCGTCACTCAGACCCATGAGCATCAGTCAAGCAG 360
DB 438 GACCAGGAGCATCAGTCAAGACCATGAGCGTCACTCAGACCCATGAGCATCAGTCAAGCAG 497
QY 361 GACATCATCTCAGCAGCATGATCATCAGTCTCAGCATATATCATGCTTCTGTTAAAAAT 420
DB 498 GACATCATCTCAGCAGCATGATCATCAGTCTCAGCATATATCATGCTTCTGTTAAAAAT 557
QY 421 AAGCGAAAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAAGTAAAGATCTTCAAGC 480
DB 558 AAGCGAAAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAAGTAAAGATCTTCAAGC 617
QY 481 AGCCAGGGGAAAGGAGCTCACCGGACAGAACATGCGAGTGGTAGAAGGAATGTCAAATGAC 540
DB 618 AGCCAGGGGAAAGGAGCTCACCGGACAGAACATGCGAGTGGTAGAAGGAATGTCAAATGAC 677

QY 541 ACTGTTAGTGTAGTGAAGTGACCTCAACTGTGTACAACTGTCTCTGAAGGAATCCAC 600
DB 678 ACTGTTAGTGTAGTGAAGTGACCTCAACTGTGTACAACTGTCTCTGAAGGAATCCAC 737
QY 601 TTTCTAGAGACAATAGAGACTCCAAAGACTGGAAAACCTTTCCCAAGAGATGTAAAGCAGC 660
DB 738 TTTCTAGAGACAATAGAGACTCCAAAGACTGGAAAACCTTTCCCAAGAGATGTAAAGCAGC 797
QY 661 TCCACTCCACCAGTGTACATCAAAAGAGCGGGTGAAGCCGGCTGGTGGTAGAAAACA 720
DB 798 TCCACTCCACCAGTGTACATCAAAAGAGCGGGTGAAGCCGGCTGGTGGTAGAAAACA 857
QY 721 AATGAATCTGTAGTGAAGCCGAAAAGGCTTTATGTATCCAGAAAACACAAAATGAAAAT 780
DB 858 AATGAATCTGTAGTGAAGCCGAAAAGGCTTTATGTATCCAGAAAACACAAAATGAAAAT 917
QY 781 CCTCAGGAGTGTTCATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAATGATCA 840
DB 918 CCTCAGGAGTGTTCATGATCAATGATCAATGATCAATGATCAATGATCAATGATCAATGATCA 977
QY 841 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCCCATCATCAACCAAAATTTGATGCT 900
DB 978 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCCCATCATCAACCAAAATTTGATGCT 1037
QY 901 AGATCTTGTCTGATTCATACAAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTTC 960
DB 1038 AGATCTTGTCTGATTCATACAAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTTC 1097
QY 961 TTCAAAATAGCCTGGGTTGGTGGTATATAGCCATTTCCATCATCATCATCATCATCATCATCAT 1020
DB 1098 TTCAAAATAGCCTGGGTTGGTGGTATATAGCCATTTCCATCATCATCATCATCATCATCATCAT 1157
QY 1021 CTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGTCTTTCAAATTTCTCCTGAGTTTC 1080
DB 1158 CTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGTCTTTCAAATTTCTCCTGAGTTTC 1217
QY 1081 CTGTGGCACTGGCGGTTGGGACTTTGAGTGGTGTGCTTTTACACCTTCTTCTCCACAT 1140
DB 1218 CTGTGGCACTGGCGGTTGGGACTTTGAGTGGTGTGCTTTTACACCTTCTTCTCCACAT 1277
QY 1141 TCTCATGCAAGTCAACAGCATAGTATAGCCATGAAGAACAGCAATGGAATGAAAGA 1200
DB 1278 TCTCATGCAAGTCAACAGCATAGTATAGCCATGAAGAACAGCAATGGAATGAAAGA 1337
QY 1201 GGACCACTTTTCAGTCACTGTCTCTCAAAACATCAAGAAAGTGCCTATTTTGGATTTC 1260
DB 1338 GGACCACTTTTCAGTCACTGTCTCTCAAAACATCAAGAAAGTGCCTATTTTGGATTTC 1397
QY 1261 ACGTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTCTTGTGAACATGTC 1320
DB 1398 ACGTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTCTTGTGAACATGTC 1457
QY 1321 CTCACATGTATCAACCAATTTAAAGATTAAGAAAAGAAATCAGAAAACCTGAAAAT 1380
DB 1458 CTCACATGTATCAACCAATTTAAAGATTAAGAAAAGAAATCAGAAAACCTGAAAAT 1517
QY 1381 GATGATGATGAGAGATTAAAGAGCAGTTGTCGAAGTATGAATCTCAACTTTTCAACCAAT 1440
DB 1518 GATGATGATGAGAGATTAAAGAGCAGTTGTCGAAGTATGAATCTCAACTTTTCAACCAAT 1577
QY 1441 GAGGAGAAAGTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1578 GAGGAGAAAGTATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1637
QY 1501 CCTTCCCACTTTGATCTCAGCAGCCTGAGTCTTGAAGAAAGAGAGGTCATGATAGCT 1560
DB 1638 CCTTCCCACTTTGATCTCAGCAGCCTGAGTCTTGAAGAAAGAGAGGTCATGATAGCT 1697
QY 1561 CATGCTCATCCACAGGAGGTCACAAATGAATATGTATCCAGAGGGTGCAGAGAAATGATGC 1620
DB 1698 CATGCTCATCCACAGGAGGTCACAAATGAATATGTATCCAGAGGGTGCAGAGAAATGATGC 1757

QY	1621	CATTACATTTTCACGATACACTCGGCCAGTCAGACGATCTCATTTCAACAGCATCATGAC	1680
Db	1758	CATTACATTTTCCACGATACACTCGGCCAGTCAGACGATCTCATTTCAACAGCATCATGAC	1817
QY	1681	TACCATCATATTCGTCATCATCATCAGCACCACCAAAACCAACCTCTCAACATCAGAGCCAG	1740
Db	1818	TACCATCATATTCGTCATCATCATCAGCACCACCAAAACCAACCTCTCAACATCAGAGCCAG	1877
QY	1741	CGTACTCTTCGGGAGGAGCTGAAGATCGCGCGCTGCGCACTTTGGCTGGATGCTGATA	1800
Db	1878	CGTACTCTTCGGGAGGAGCTGAAGATCGCGCGCTGCGCACTTTGGCTGGATGCTGATA	1937
QY	1801	ATGCGTATGAGGCTGTCACAAATTTTCAGCGATGGCCCTAGCAATTTGGTGTCTTTTACTCAA	1860
Db	1938	ATGCGTATGAGGCTGTCACAAATTTTCAGCGATGGCCCTAGCAATTTGGTGTCTTTTACTCAA	1997
QY	1861	GGCTTATCAAGTGGTTTAAAGTACTCTCTGTGTGTCTGTCTGTCTGTCTGTCTTTAT	1920
Db	1998	GGCTTATCAAGTGGTTTAAAGTACTCTGTGTGTGTCTGTCTGTCTGTCTGTCTTTAT	2057
QY	1921	TTAGGTGACCTTTCCTCTTCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGTCTTTAT	1980
Db	2058	TTAGGTGACCTTTCCTCTTCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGTCTTTAT	2117
QY	1981	ANTGCATTTGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAAATTTTCATTTGGTCAT	2040
Db	2118	ANTGCATTTGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAAATTTTCATTTGGTCAT	2177
QY	2041	TATGCTGAAATTTTCTATGTGGATATTTTGACCTTACTGCTGGCTTATATGATGATGTT	2100
Db	2178	TATGCTGAAATTTTCTATGTGGATATTTTGACCTTACTGCTGGCTTATATGATGATGTT	2237
QY	2101	GTCTCTGTTTATATGCTTACCTGAAATGCTGCACAATGATGCTACTGACCATGGATGAGC	2160
Db	2238	GTCTCTGTTTATATGCTTACCTGAAATGCTGCACAATGATGCTACTGACCATGGATGAGC	2297
QY	2161	CGCTGGGGGATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATTTATGCTTA	2220
Db	2298	CGCTGGGGGATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATTTATGCTTA	2357
QY	2221	CTTATTTCCATATTTGAACATAAAATCTGTTTCTGATATAAATTTCTAG	2268
Db	2358	CTTATTTCCATATTTGAACATAAAATCTGTTTCTGATATAAATTTCTAG	2405
RESULT 2			
AHA92299			
Db	AHA92299 standard; cDNA; 3461 BP.		
XX	AHA92299;		
Ac			
XX	10-JUN-2002 (first entry)		
DE	Human breast cancer 4 gene (BCR4) cDNA.		
XX			
KW	BCR4; human; breast cancer 4 gene; prostate cancer; diagnosis;		
KW	gene therapy; vaccine; biochip; chromosome 18q12; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
FI	Key	Location/Qualifiers	
FT	ChS	138..2405	
FT	/*tag= a		
FT	/*product= "BCR4"		
FT	sig_peptide	138..197	
FT	/*tag= b		
FT	mat_peptide	198..2402	
FT	/*tag= c		
XX			
PN	WU200216939-A2.		
XX			
PD	28-FEB-2002.		
XX			

PF	20-AUG-2001; 2001WO-US25997.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
----	------------------------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

bb	498	GAGCATCACTCTGGACCATGATCAACACATCTACCACTAATCATGCTGCTTCTGGTAAAAAT	557
OY	421	AAGCTGAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAGAGATCTTAGAAC	480
bb	558	AAGCTGAAAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAGAGATCTTAGAAC	617
OY	481	AGCCAGGCGAAGGAGCTCACCGACGAGCAATGCCAGTGGTAGAAGGAATCTCAAGAC	540
bb	618	AGCCAGGCGAAGGAGCTCACCGACGAGCAATGCCAGTGGTAGAAGGAATCTCAAGAC	677
OY	541	ACTGTATTAGTGTACTGGAAGTGACCTCAACTGTGTACAACTACTGCTCTCTCAAGCAATCTAC	600
bb	678	ACTGTATTAGTGTACTGGAAGTGACCTCAACTGTGTACAACTACTGCTCTCTCAAGCAATCTAC	737
OY	601	TTTTCTAGAGACAATAGAGACTTCCAAGACCTGGAAACATCTTCCCCCAAGATCTTAAGTAGC	660
bb	738	TTTTCTAGAGACAATAGAGACTTCCAAGACCTGGAAACATCTTCCCCCAAGATCTTAAGTAGC	797
OY	661	TCCACTCCACCGAGTGTACATCAAAAGACCGGGTGAGTCGGCTGGCTAGGAAACAC	721
bb	798	TCCACTCCACCGAGTGTACATCAAAAGACCGGGTGAGTCGGCTGGCTAGGAAACAC	857
OY	721	AATCAATCTGTGACTGAGCCCCCAAAAGGCTTTATGTATTCAGAAACACAAATTAAGAT	780
bb	858	AATCAATCTGTGACTGAGCCCCCAAAAGGCTTTATGTATTCAGAAACACAAATTAAGAT	917
OY	781	CTTCAGGAGTGTTTTCAATGCATCAAAAGCTACTGACATCTCATGGCATGGGCATTCAGGTT	840
bb	918	CTTCAGGAGTGTTTTCAATGCATCAAAAGCTACTGACATCTCATGGCATGGGCATTCAGGTT	977
OY	841	CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAACCAAAATTCATGCT	900
bb	978	CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAACCAAAATTCATGCT	1037
OY	901	AGATCTTGTCTGTGATTCATCAACTGAAAAGAGGCTGAAATCCCTCCCAACACCTATATCA	960
bb	1038	AGATCTTGTCTGTGATTCATCAACTGAAAAGAGGCTGAAATCCCTCCCAACACCTATATCA	1097
OY	961	TTTACAATACCCCTGGGTGGTGGTTTTATAGGCATTTCCATCATGACTTTCTCTCTCTCTG	1020
bb	1098	TTTACAATACCCCTGGGTGGTGGTTTTATAGGCATTTCCATCATGACTTTCTCTCTCTCTG	1157
OY	1021	CTGGGCGCTTATCTTAGTGGCTCTCATGAATCGGGTGTTTTTCAAAATTCCTCTCAGTTTC	1080
bb	1158	CTGGGCGCTTATCTTAGTGGCTCTCATGAATCGGGTGTTTTTCAAAATTCCTCTCAGTTTC	1217
OY	1081	CTTGTGGCACTGGCCCTTGGGACTTTTGAGTGGTGATGCTTTTTTACACCTCTCTTCCACAT	1140
bb	1218	CTTGTGGCACTGGCCCTTGGGACTTTTGAGTGGTGATGCTTTTTTACACCTCTCTTCCACAT	1277
OY	1141	TCTCATGCCAGTCAACACCATATGCTATAGCCATGAAGAACGAGCAATGGAAATGAAAGA	1200
bb	1278	TCTCATGCCAGTCAACACCATATGCTATAGCCATGAAGAACGAGCAATGGAAATGAAAGA	1337
OY	1201	GGACCTACTTTTTCAGTTCATCTGTCTTCTCAAAACATAGAGAAGAGTGCCTATTCTGATTCG	1260
bb	1338	GGACCTACTTTTTCAGTTCATCTGTCTTCTCAAAACATAGAGAAGAGTGCCTATTCTGATTCG	1397
OY	1261	ACGTGGAGCGTCTTAACAGCTCTAGAGCGCTGTATTTCATGTTCTTGTGTGAACATGTC	1320
bb	1398	ACGTGGAGCGTCTTAACAGCTCTAGAGCGCTGTATTTCATGTTCTTGTGTGAACATGTC	1457
OY	1321	CTCACATTCATCAACCAATTTAAACATAGAAGAAAGAAATCAGAACCAAACTGCAAAAT	1380
bb	1458	CTCACATTCATCAACCAATTTAAACATAGAAGAAAGAAATCAGAACCAAACTGCAAAAT	1517
OY	1381	GATGATGATGTGGAGATTAAAGAGCAGTGTGTCGAAGTATGAATCTCAACTTTTCAACAAAT	1440
bb	1518	GATGATGATGTGGAGATTAAAGAGCAGTGTGTCGAAGTATGAATCTCAACTTTTCAACAAAT	1577
OY	1441	GAGCAAAAGTATAGATACAGATGATCGAACTGAAGGCTATTTTACGAGCAGAACTTACAAAG	1500

Db	1578	GAGGAGAAAGTAGATACAGATGATCGAACTCGAAGGCGTATTTACGAGCAGACTCAACAG	16137
QY	1501	CCCTCCCACTTTGATTCCTCAGCAGCCTGCAGTCTCGGAAGAAGAGGTGATCATGATAGCT	1560
Db	1638	CCCTCCCACTTTGATTCCTCAGCAGCCTGCAGTCTTGGAAAGAAGAAGGTGATCATGATAGCT	1697
QY	1561	CATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCACAGAGGTGCAAGAATAAATGC	1620
Db	1698	CATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCACAGAGGTGCAAGAATAAATGC	1757
QY	1621	CATTTCACATTTCCACGATACACTCGCCAGTCAGAGGATCTCATTTACCACCACCATCATGAC	1680
Db	1758	CATTTCACATTTCCACGATACACTCGCCAGTCAGAGGATCTCATTTACCACCACCATCATGAC	1817
QY	1681	TACCATCATATTTCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	1740
Db	1818	TACCATCATATTTCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC	1877
QY	1741	CGCTACTCTCGGAGGAGCTGAAGATCGCGGCTGCGCACTTTGGCCCTGGATGGTGATA	1800
Db	1878	CGCTACTCTCGGAGGAGCTGAAGATCGCGGCTGCGCACTTTGGCCCTGGATGGTGATA	1937
QY	1801	ATGGGTGATGGCCTGGCACATTTTCACGGATGGCCTAGCAATTTGGTGCCTTTTACTGAA	1860
Db	1938	ATGGGTGATGGCCTGGCACATTTTCACGGATGGCCTAGCAATTTGGTGCCTTTTACTGAA	1997
QY	1861	GGCTTATCAAGTGGTTTAAGTACTCTGTGGCTGTGTCTGTGCATGAGTTGGCCTCATGAA	1920
Db	1998	GGCTTATCAAGTGGTTTAAGTACTCTGTGGCTGTGTCTGTGCATGAGTTGGCCTCATGAA	2057
QY	1921	TTAGTGACATTTGCTGTCTACTAAAGGCTGGCAGTACCGTTAAGCAGGCTGTGCTTTAT	1980
Db	2058	TTAGTGACATTTGCTGTCTACTAAAGGCTGGCAGTACCGTTAAGCAGGCTGTGCTTTAT	2117
QY	1981	AATGCATTGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTTCATTGGTCTAT	2040
Db	2118	AATGCATTGTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTTCATTGGTCTAT	2177
QY	2041	TATGCTGAAAATGTTTCTATGTGGATATTGCGACTTACTGCTGGCTTTTTCATTGGTCTAT	2100
Db	2178	TATGCTGAAAATGTTTCTATGTGGATATTGCGACTTACTGCTGGCTTTTTCATTGGTCTAT	2237
QY	2101	GCTCTGGTTGATATGTTAGCTGAAATGCTGCACATGCTAGTGACCATGGATGTAGC	2160
Db	2238	GCTCTGGTTGATATGTTAGCTGAAATGCTGCACATGCTAGTGACCATGGATGTAGC	2297
QY	2161	CGCTGGGGGTATTTCTTTTACAGAACTGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2220
Db	2298	CGCTGGGGGTATTTCTTTTACAGAACTGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2357
QY	2221	CTTATTTCCATATTGAACATAAAATCGTGTTCGTATAAATTTCTAG	2268
Db	2358	CTTATTTCCATATTGAACATAAAATCGTGTTCGTATAAATTTCTAG	2405
RESULT 3			
AAD13480			
ID	AAD13480 standard; cdna; 2776 bp.		
AC	AAD13480;		
CC			
DT	06-NOV-2001 (first entry)		
DE	Human LIV-1-164647 cdna.		
KW	Human; LIV-1-164647; cytosstatic; estrogen-inducible gene; tumour;		
KW	cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;		
KW	salivary gland; carcinoma; drug screening; therapy; ss.		
OS	Homo sapiens.		
Key	Local/Qualifiers		
FT	73-2340		
CDS			

FT	/+tag= a
FT	/product= "Human LIV-1-164647 protein"
FN	WO200155178-A2.
PN	02-AUG-2001.
PI	25-JAN-2001; 2001WO-US02622.
PP	25-JAN-2000; 2000US-0177951.
PR	10-APR-2000; 2000US-0195761.
XX	(GPTH) GENENTECH INC.
PA	Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
XX	WPI: 2001 502628/55.
XX	P-PSDB; AAK07206.
XX	New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
PT	the proliferation of tumor cells in mammals, e.g. breast, lung,
PT	prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
PT	.
XX	Claim 6; Fig 2A; 150pp; English.
US	The present sequence is human estrogen-inducible LIV-1-164647
XX	cDNA. LIV-1 is overexpressed in tumour tissues such as prostate,
CC	colon, lung and breast. The LIV-1 DNA and polypeptide, and the monoclonal
CC	antibody are useful for treating cancer and inhibiting the proliferation
CC	of tumour cells in mammals, e.g. breast, lung, prostate, colon, ovary,
CC	uterus, kidney, gastric or salivary carcinoma, or other tumour cell
CC	types expressing the LIV-1-164647 protein. In particular, the mammal is
CC	a human. The LIV-1 DNA and polypeptide may also be used in screening
CC	assays for drug candidates.
XX	Sequence 2776 BP; 823 A; 583 C; 592 G; 778 T; 0 other;
XX	
Query Match	99.9%; Score 2264.8; DB 22: Length 2776;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2266;	Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGCGGAGGAAGTTATCTGTAATCTTTCATCTGACCTTTGCCCTCTCTGTACAAATGCC 60
DB	73 A'GCGGAGGAAGTTATCTGTAATCTGACCTTTGCCCTCTCTGTACAAATGCC 132
QY	61 CTTCATCAACTAAAGCAGCTTGCTTCCCGCAGACCACTGAGAAATTTAGTCCAAATTGG 120
DB	133 CTTCATCAAGCAAGCAGCTTGCTTCCCGCAGACCACTGAGAAATTTAGTCCAAATTGG 192
QY	121 GAATCTGGCATTAATGTTGACTTGGCAATTTCCACGGCAATATCATCTACAACAGCTT 180
DB	193 GAATCTGGCATTAATGTTGACTTGGCAATTTCCACGGCAATATCATCTACAACAGCTT 252
QY	181 TTCTACCGCTATGAGAAAAATAATCTTTGTGACTTTGAAGGGTTTCAGAAAATTTACTTTCAA 240
DB	253 TTCTACCGCTATGAGAAAAATAATCTTTGTGACTTTGAAGGGTTTCAGAAAATTTACTTTCAA 312
QY	241 AATATAGCATAGATAAGATTTAAAGAAATCCATATACCATGACCACGACCATCACTCA 300
DB	313 AATATAGCATAGATAAGATTTAAAGAAATCCATATACCATGACCACGACCATCACTCA 372
QY	301 GACACAGGATCACTCAGACCATGAGCGTCACTTCAGACCATGAGCATCACTCAGACGAC 360
DB	173 GACCACAGCATCACTCAGACCATGAGCGTCACTTCAGACCATGAGCATCACTCAGACGAC 432
QY	361 GAGCATCACTCTGACCATGATCATCACTTCACCATTAATCATGCTGCTTCGTGTAATAAT 420
DB	433 GAGCATCACTCTGACCATGATCATCACTTCACCATTAATCATGCTGCTTCGTGTAATAAT 492
QY	421 AAGCGAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGTAAAGATCTTAGAAC 480
DB	493 AAGCGAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGTAAAGATCTTAGAAC 552

PR	09-JUN-2000; 2000US-211J14P.
PR	18-JUL-2000; 2000US-219007P.
PR	13-DEC-2000; 2000US-255281P.
XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	Schlegel R, Endege WO, Monahan JE;
PJ	WPI; 2001-662795/76.
DR	
XX	
PT	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer -
PT	
XX	
PS	Claim 1; Page 5169-5170; 11750pp; English.
XX	
CC	The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABY00010-ABV62213) of the specification or its complement. (1) is useful for:
CC	(a) assessing whether a patient is afflicted with prostate cancer;
CC	(b) monitoring the progression of prostate cancer in a patient;
CC	(c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;
CC	(d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;
CC	(f) assessing the prostate cell carcinogenic potential of a compound;
CC	(g) determining whether prostate cancer has metastasized in a patient;
CC	(h) assessing the aggressiveness or indolence of prostate cancer in a patient;
CC	(i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX	
QQ	Sequence 3523 BP; 1071 A; 684 C; 735 G; 1013 T; 20 other;
Query Match	97.9%; Score 2220.4; DB 23; Length 3523;
Best Local Similarity	99.2%; Pred. No. 0;
Matches 2249; Conservative	0; Mismatches 1; Indels 18; Gaps 1;
Qy	1 ATGCGAGGAAGTTATCTGTAAATCTTGATCTGACCTTGCCTCTCTGTACAAAATCCC 60
Db	138 ATGCGAGGAAGTTATCTGTAAATCTTGATCTGACCTTGCCTCTCTGTACAAAATCCC 197
Qy	61 TTTCATGAACATAAAGCAGCTGCTTCCCCAGACCACACTGAGAAAAATTAGTCCGAATTGG 120
Db	198 TTTCATGAACATAAAGCAGCTGCTTCCCCAGACCACACTGAGAAAAATTAGTCCGAATTGG 257
Qy	121 GAATCTGGCATTAATGTTGACTTGGCAATTTCCACACGGCAATATCATCTACACACGCTT 180
Db	258 GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGGCAATATCATCTACACACGCTT 317
Qy	181 TTCTACCGCTATGAGAAAAATAATCTTTGTCACTTCAAGGGTTACAAAAATTACTTCAA 240
Db	318 TTCTACCGCTATGAGAAAAATAATCTTTGTCACTTCAAGGGTTACAAAAATTACTTCAA 377
Qy	241 AATATAGGCATPAGATAAGATTTAAAGAATCCATATACACCATGACCCAGCACCATCACTCA 300
Db	378 AATATAGGCATAGATAAGATTTAAAGAATCCATATACACCATGACCCAGCACCATCACTCA 437
Qy	301 GACCACGAGCATCACTCAGACCATGACCGTCACTCAGACCATGAGCATCACTCAGACCCAC 360
Db	438 GACCACGAGCATCACTCAGACCATGACCGTCACTCAGACCATGAGCATCACTCAGACCCAC 497
Qy	361 GAGCATCACTCTGACCATGATCATCTCTCACCATAATCATGCTGCTCTCGTAAAAAT 420
Db	498 GAGCATCACTCTG-----ACCATANTCATGCTGCTCTCGTAAAAAT 539
Qy	421 AAGCGAAAGCTCTTTGCCCCAGCACCATGACTCAGATAGTTTCAGTAAAGATTCCTAGAAAC 480
Db	540 AAUCGAAAGCTCTTTGCCCCAGCACCATGACTCAGATAGTTTCAGTAAAGATTCCTAGAAAC 599
Qy	481 AGCCAGGGGAAAGAGCTCACCCAGACAATGCCAGTGGTAGAGGAATGTCAAGGAC 540
Db	600 AGCCAGGGGAAAGAGCTCACCCAGACAATGCCAGTGGTAGAGGAATGTCAAGGAC 659

QY 541 AGTGTACTGCTAGTGAAGTCACCTCAACTGCTGTACACACTGTCCTGAAGCAACTCAC 600
Db 660 ACTGTTAGTGTAGTGAAGTCACCTCAACTGCTGTACACACTGTCCTGAAGCAACTCAC 719
QY 601 TTTCFAGAGACAATAGAGACTCCAGAGCTTGGAAACCTCTTCCCAAGATGTAAGCAGC 660
Db 720 TTTCFAGAGACAATAGAGACTCCAGAGCTTGGAAACCTCTTCCCAAGATGTAAGCAGC 779
QY 661 TCCACTCCACCACCTGTACATCAAGAGCGGGTGAGCGGCTGGCTGTGTAGGAAACA 720
Db 780 TCCACTCCACCACCTGTACATCAAGAGCGGGTGAGCGGCTGGCTGTGTAGGAAACA 839
QY 721 AATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 780
Db 840 AATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 899
QY 781 CCTCAGAGTGTTCATGCAATCAAGCTACGTACATCTCATGGCATGGGCATCCAGGTT 840
Db 900 CCTCAGAGTGTTCATGCAATCAAGCTACGTACATCTCATGGCATGGGCATCCAGGTT 959
QY 841 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCATCAACCAAAATGATGCT 900
Db 960 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCATCAACCAAAATGATGCT 1019
QY 901 AGATCTTGTCTGATTCATACAGTGAAGAGAGGCTGAAATCCCTCCAAAGACCTATTCA 960
Db 1020 AGATCTTGTCTGATTCATACAGTGAAGAGAGGCTGAAATCCCTCCAAAGACCTATTCA 1079
QY 961 TTCAAAATAGCCGCTGGTGGTGTATAGCCATTTCCATCATCAGTTTCCCTGCTCTG 1020
Db 1080 TTCAAAATAGCCGCTGGTGGTGTATAGCCATTTCCATCATCAGTTTCCCTGCTCTG 1139
QY 1021 TCGGGGCTATCTTAGTGCCTCTCATGAATCGGGTGTATTTTCAAAATTTCCCTGACTTTC 1080
Db 1140 CTGGGGCTATCTTAGTGCCTCTCATGAATCGGGTGTATTTTCAAAATTTCCCTGAGTTTC 1199
QY 1081 CTCTGGCACTGGCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTCTCTCCACAT 1140
Db 1200 CTCTGGCACTGGCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTCTCTCCACAT 1259
QY 1141 TCTCATGCAAGTCACCACTAGTCATAGCCATGAAGAACCCAGCAATGCAAAATCAAAAGA 1200
Db 1260 TCTCATGCAAGTCACCACTAGTCATAGCCATGAAGAACCCAGCAATGCAAAATCAAAAGA 1319
QY 1201 GGACCACTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAAAGTGCCTATTTTCATTC 1260
Db 1320 GGACCACTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAAAGTGCCTATTTTCATTC 1379
QY 1361 ACCTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTGTGAACATGTC 1320
Db 1380 ACCTGGAAGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTGTGAACATGTC 1439
QY 1321 CTCACATTGATCAACAAATTTAAAGATAAGAGAAAAGAAATCAGAGAACTCAAAAT 1380
Db 1440 CTCACATTGATCAACAAATTTAAAGATAAGAGAAAAGAAATCAGAGAACTCAAAAT 1499
QY 1381 GATGATGATGAGATTAAGAGAGAGTGTCCAAAGTATGAATCTCAACTTTCAACAAAT 1440
Db 1500 GATGATGATGAGATTAAGAGAGAGTGTCCAAAGTATGAATCTCAACTTTCAACAAAT 1559
QY 1441 GAGGAGAACTAGATACAGATGATCAACTGAAGGCTATTTACGACGACACTCACAGAG 1500
Db 1560 GAGGAGAACTAGATACAGATGATCAACTGAAGGCTATTTACGACGACACTCACAGAG 1619
QY 1501 CCCTCCCACTTTGATTTCTCAGCAGGCTGCAGTCTTGGAAAGAGAGAGGTCATGATGCT 1560
Db 1620 CCCTCCCACTTTGATTTCTCAGCAGGCTGCAGTCTTGGAAAGAGAGAGGTCATGATGCT 1679
QY 1561 CATGCTCATCCAGAGAACTCTACAATGAATATGTACCCAGAGGCTGCAAGAAATAAATGC 1620
Db 1680 CATGCTCATCCAGAGAACTCTACAATGAATATGTACCCAGAGGCTGCAAGAAATAAATGC 1739

QY 1621 CATTCACATTTCCAGATACACTCGGCCAGTCAGACGATCTCATTCACCACCATCATGAC 1680
Db 1740 CATTCACATTTCCAGATACACTCGGCCAGTCAGACGATCTCATTCACCACCATCATGAC 1799
QY 1681 TACCATCATATCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1740
Db 1800 TACCATCATATCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1859
QY 1741 CGCTACTCTCGGAGGAGCTGAAAGATGCGCGCTGCCCACTTTGGCCTGGATGGTGATA 1800
Db 1860 CGCTACTCTCGGAGGAGCTGAAAGATGCGCGCTGCCCACTTTGGCCTGGATGGTGATA 1919
QY 1801 ATGGGTGATGGCTGCACAAATTTACAGGATGGCTAGCAATTTGGTGTCTCTTTACTGAA 1860
Db 1920 ATGGGTGATGGCTGCACAAATTTACAGGATGGCTAGCAATTTGGTGTCTCTTTACTGAA 1979
QY 1861 GGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGT 1920
Db 1980 GGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGT 2039
QY 1921 TTAGTGACTTTGCTGTCTTACTAAAGGCTGGCATGACCGTTTAAGCAGGCTCTCTTTAT 1980
Db 2040 TTAGTGACTTTGCTGTCTTACTAAAGGCTGGCATGACCGTTTAAGCAGGCTCTCTTTAT 2099
QY 1981 AATGCATTTGTCAAGCATGCTGGGTATCTTGAATGGCAACAGAGAAATTTTCATTTGTCAT 2040
Db 2100 AATGCATTTGTCAAGCATGCTGGGTATCTTGAATGGCAACAGAGAAATTTTCATTTGTCAT 2159
QY 2041 TATGCTGAAATTTTCTTATGTGGATATTTGCACCTTACTGCTGGCTTATTCATGTATGTT 2100
Db 2160 TATGCTGAAATTTTCTTATGTGGATATTTGCACCTTACTGCTGGCTTATTCATGTATGTT 2219
QY 2101 GCTCTGGTGATATGGTACCTGAAATGCTGCAATGATGCTAGTGACCATGGATGTAGC 2160
Db 2220 GCTCTGGTGATATGGTACCTGAAATGCTGCAATGATGCTAGTGACCATGGATGTAGC 2279
QY 2161 CGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATTAATGTTA 2220
Db 2280 CGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATTAATGTTA 2339
QY 2221 CTTATTTCCATATTTGACATAAAATCGTGTTCGTATATAATTTCTAG 2268
Db 2340 CTTATTTCCATATTTGACATAAAATCGTGTTCGTATATAATTTCTAG 2387

RESULT 5
AAD13479 standard; DNA; 3461 BP.
XX AAD13479;
XX AC
XX DT 06-NOV-2001 (first entry)
XX DE Human LIV-1 gene.
XX KW Human; LIV-1; cytostatic; estrogen-inducible gene; tumour; therapy;
KW cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;
KW salivary gland; carcinoma; drug screening; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 138..2396
FT FT /*tag= a
FT FT /product= "Human LIV-1 protein"
FT FT 138..197
FT FT /*tag= b
FT FT 198..2393
FT FT /*tag= c
FT FT /product= "Mature LIV-1 protein"
XX PN WO200155178-A2.
XX

PD 02-AUG-2001.
 XX 25-JAN-2001: 2001WO-US02622.
 XX 25-JAN-2000: 2000US-0177951.
 PR 10-APR-2000: 2000US-0195761.
 XX (GENE) GENENTECH INC.
 PA Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
 PJ WPL: 2001-502628/55.
 XX P-PSDB: AAE07205.
 XX New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
 PT the proliferation of tumor cells in mammals, e.g. breast, lung,
 PT prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
 PT
 XX Example 2; Fig 1A; 150pp; English.
 PS
 XX The present sequence is human estrogen-inducible gene, (LIV-1).
 CC LIV-1 is overexpressed in tumor tissues such as prostate, colon,
 CC lung and breast. The LIV-1 DNA and polypeptide, and the monoclonal
 CC antibody are useful for treating cancer and inhibiting the proliferation
 CC of tumor cells in mammals, e.g. breast, lung, prostate, colon, ovary,
 CC uterus, kidney, gastric or salivary carcinoma, or other tumor cell
 CC types expressing the LIV-1-164647 protein. In particular, the mammal is
 CC a human. The LIV-1 DNA and polypeptide may also be used in screening
 CC assays for drug candidates.
 XX
 XX Sequence 3461 BP; 1057 A; 679 C; 727 G; 998 T; 0 other;
 SQ
 Query Match 96.9%; Score 2198; DB 22; Length 1461;
 Best local similarity 99.1%; Prod. No. 0;
 Mismatches 2248; Conservative 0; Mismatches 0; Indels 20; Gaps 3;
 QY 1 ATGGGAGAGAAAGTATCTGTATCTGATCTGACCTTGGCCCTCTCTGTCAAAATGCC 60
 DB 118 ATGGGAGAGAAAGTATCTGTATCTGATCTGACCTTGGCCCTCTCTGTCAAAATGCC 197
 QY 61 CTTCATGCAAGTAAAGAGAGTCTTTTCCCGAGACCTGAGAAATTTAGTCCGAATTCG 120
 DB 198 CTTCATGCAAGTAAAGAGAGTCTTTTCCCGAGACCTGAGAAATTTAGTCCGAATTCG 257
 QY 121 GAATCTGCAATTAATTTGACCTTGCAATTTCCACAGCGCAATATCATCTACACAGCTT 180
 DB 258 GAATCTGCAATTAATTTGACCTTGCAATTTCCACAGCGCAATATCATCTACACAGCTT 317
 QY 181 TTCTACCTGATGAGAAATTAATTTGTCAGTTGAAGGTTTCAGAAATTTACTTCAA 240
 DB 318 TTCTACCTGATGAGAAATTAATTTGTCAGTTGAAGGTTTCAGAAATTTACTTCAA 377
 QY 241 AATATAGGCATAGATTAAGATTTAAAGAAATCCATATACACCATGACGACCATCACATCA 300
 DB 378 AATATAGGCATAGATTAAGATTTAAAGAAATCCATATACACCATGACGACCATCACATCA 437
 QY 301 GACACGAGCATCTACAGCATGAGCGTCACTACAGACCATGACATCACTACAGACCAT 360
 DB 418 GACACGAGCATCTACAGCATGAGCGTCACTACAGACCATGACATCACTACAGACCAT 497
 QY 361 GAGCATCACTCTGACCATGATCATCACTCTCCACCATATCATCTGCTTCTGTAATAAT 420
 DB 498 GAGCATCACTCTG-----ACCATATCATCTGCTTCTGTAATAAT 539
 QY 421 AAGTGAAAGGCTTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTACGAAC 480
 DB 540 AAGTGAAAGGCTTTTGGCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTACGAAC 549
 QY 481 AGCCAGGGAAGGAGCTTACCGACCGACGACCATGCCACTGCTGTAAGAAATTTTAAAGAC 540
 DB 600 AGCCAGGGAAGGAGCTTACCGACCGACGACCATGCCACTGCTGTAAGAAATTTTAAAGAC 659
 QY 541 AGTCTTAGTCTAGTGAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 600
 DB 660 AGTCTTAGTCTAGTGAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 719
 QY 601 TTTCTAGAGACAATAGAGACTCCAAGACCTCGAAAACCTTTTCCCAAGATGTAAAGACG 660
 DB 720 TTTCTAGAGACAATAGAGACTCCAAGACCTCGAAAACCTTTTCCCAAGATGTAAAGACG 779
 QY 661 TCCACTCCACCCAGTGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 720
 DB 780 TCCACTCCACCCAGTGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 839
 QY 721 AATGAATCTGTAGTGAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 780
 DB 840 AATGAATCTGTAGTGAAGTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 899
 QY 781 CCTCAGGAGTCTTTCAATGCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 840
 DB 900 CCTCAGGAGTCTTTCAATGCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 959
 QY 841 CCGCTGAATGCAACAGAGTTCACATCTCTGTCCAGCCATCATCACTCACTCACTCACTCACTCA 900
 DB 960 CCGCTGAATGCAACAGAGTTCACATCTCTGTCCAGCCATCATCACTCACTCACTCACTCACTCA 1019
 QY 901 AGATCTTGTCTGTATCATCAAGTGAAGAGAGCTCAAAATCCCTCCAAAGACCTATTTCA 960
 DB 1020 AGATCTTGTCTGTATCATCAAGTGAAGAGAGCTCAAAATCCCTCCAAAGACCTATTTCA 1079
 QY 961 TTACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020
 DB 1080 TTACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1139
 QY 1021 CTGGGGCTTATCTTAGTGCCTCTCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 1080
 DB 1140 CTGGGGCTTATCTTAGTGCCTCTCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 1199
 QY 1081 CTGTGGCAGCTGGCGCTGGGAGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1140
 DB 1200 CTGTGGCAGCTGGCGCTGGGAGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1259
 QY 1141 TCTCATGCAAGTCAACACCATAGTCTAGTCCATGAAGAACCAAGCAATGGAATGGAAGA 1200
 DB 1260 TCTCATGCAAGTCAACACCATAGTCTAGTCCATGAAGAACCAAGCAATGGAATGGAAGA 1319
 QY 1201 GGACCACTTTTCACTCATCTCTCTCAAAACATAGAAAGAGTGCCTATTTTGGATTCC 1260
 DB 1320 GGACCACTTTTCACTCATCTCTCTCAAAACATAGAAAGAGTGCCTATTTTGGATTCC 1379
 QY 1261 ACCTGGAAGGCTCTAACAGCTCTAGGAGGCTCTATTTTCACTGTTTCTTGTGAACATGTC 1320
 DB 1380 ACCTGGAAGGCTCTAACAGCTCTAGGAGGCTCTATTTTCACTGTTTCTTGTGAACATGTC 1439
 QY 1321 CTCACTTTGATCAACCAATTTAAAGATAGAAAGAAAGAAATCAGAGAAACCTGAAAT 1380
 DB 1440 CTCACTTTGATCAACCAATTTAAAGATAGAAAGAAAGAAATCAGAGAAACCTGAAAT 1499
 QY 1381 GATGATGATGCGAGATTTAAGAGAGAGTGTCTCAAGTATGAAATCTCAACTTTTCAACAAAT 1440
 DB 1500 GATGATGATGCGAGATTTAAGAGAGAGTGTCTCAAGTATGAAATCTCAACTTTTCAACAAAT 1559
 QY 1441 GAGGAGAAAGTATGATACAGATGATCGAACTGAAAGGCTATTTACGAGCAGACATCACAAGAG 1500
 DB 1560 GAGGAGAAAGTATGATACAGATGATCGAACTGAAAGGCTATTTACGAGCAGACATCACAAGAG 1619
 QY 1501 CCTCTCCCACTTTGATCTCAGCAGCCTGAGCTCTGGAAGAGAGAGAGTCTATGATAGCT 1560
 DB 1620 CCTCTCCCACTTTGATCTCAGCAGCCTGAGCTCTGGAAGAGAGAGAGTCTATGATAGCT 1679
 QY 1561 CATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAGAAATTAATGCC 1620
 DB 1680 CATGCTCATCCACAGGAAGTCTACAATGAATATGTACCCAGAGGGTGCAGAAATTAATGCC 1739
 QY 1621 CATTCACATTTTCCACGATACACTCGGCCAGTCAGAGGATCTTCATTCACCAACCATCATGAC 1680

601 TACAATGAATATGTACCCAGAGGCTGCAAGAATAAATGCCATTCACATTTCCACGATACA 660
1642 CTCGGCCAGTCACAGCATCTCATTTACACCATCATGACTACCATCATATTTCCATCAT 1701
661 CTCGGCCAGTCACAGCATCTCATTTACACCATCATGACTACCATCATATTTCCATCAT 720
1702 CACACCAACCAACACCATCTCTACAGTTCACAGCCAGCGTACTCTCGGAGAGAGCTG 1761
721 CACACCAACCAACACCATCTCTACAGTTCACAGCCAGCGTACTCTCGGAGAGAGCTG 780
1762 AAGATGCCGCGCTGCCACCTTTGGCTGGATGATGATGATGATGATGATGATGATGAT 1821
781 AAGATGCCGCGCTGCCACCTTTGGCTGGATGATGATGATGATGATGATGATGATGAT 840
1822 TTCAGCGATGGCTTAGCAATTTGGCTGGCTTTTACTGAAGCTTATCAAGTGTGTTAACT 1881
841 TTCAGCGATGGCTTAGCAATTTGGCTGGCTTTTACTGAAGCTTATCAAGTGTGTTAACT 900
1882 ACTTCTGCTGCTGTTCTGCTCATGAGTTGCCCTCATGAATAGGTGACTTTGCTGTTCTA 1941
901 ACTTCTGCTGCTGTTCTGCTCATGAGTTGCCCTCATGAATAGGTGACTTTGCTGTTCTA 960
1942 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGAT 2001
961 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGAT 2020
2002 GCGTATCTTTGGAATGCCACAGGAATTTTCATTTGGTTCATGCTGAAATCTTTCTATG 2061
1021 GCGTATCTTTGGAATGCCACAGGAATTTTCATTTGGTTCATGCTGAAATCTTTCTATG 2080
2062 TGAATATTTGCACTTACTGCTGCTGCTTATTCATGATGATGATGATGATGATGATGAT 2121
1081 TGAATATTTGCACTTACTGCTGCTGCTTATTCATGATGATGATGATGATGATGATGAT 2140
2122 GAAATGCTGCACAAATGATGCTAGTACCATTGGATGCTAGCGCTGGGGTATTTCTTTTA 2181
1141 GAAATGCTGCACAAATGATGCTAGTACCATTGGATGCTAGCGCTGGGGTATTTCTTTTA 2200
2182 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTCAACAT 2241
1201 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTATTTCCATATTTCAACAT 2259
2242 AAAATGCTGTTTCGTATAAATTTCTAG 2268
1260 AAAATGCTG-TTCGTATAAATTTCTAG 1285

RESULT 7
AA199071
XX
AC AA199071 standard; cDNA: 2404 BP.
XX
XX AA199071:
XX
XX 24-MAR-1998 (first entry)
XX
XX
XX Partial sequence of the pLIV1 gene.
XX
XX Oestrogen regulated gene; pLIV1 gene partial sequence; breast cancer;
KW metastatic risk assessment; hormone therapy responsiveness;
KW tumour oestrogen receptor expression; lymph node involvement; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT 1..1296
FT CDS /*tag= a
XX
XX US5693465-A.
XX
XX 02-DEC-1997.
XX
XX 22-SEP-1994; 94US-0311023.

22-SEP-1994; 94US-0311023.
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX
XX Gee JMW, Green CD, Manning DL, Nicholson RI:
XX
XX WPI; 1998-031740/03.
XX P-PSDB; AAW34528.
XX
XX Assessment of metastatic risk or oestrogen responsiveness in breast
XX cancer - by detecting expression of pLIV1 protein
XX
XX Disclosure; Column 11-14; 11pp; English.
XX
XX This sequence represents a partial sequence of the pLIV1 gene. This
XX sequence extends further in the 3' direction than the pLIV1 partial
XX sequence shown in AAT99070. The protein encoded by this sequence is an
XX oestrogen regulated gene associated with breast cancer. The protein
XX encoded by this sequence can be used in the method of the invention. The
XX method is for determining the risk of metastasis of a female breast
XX tumour and predicting the responsiveness of a female breast tumour to
XX hormone treatment comprising determining if a tissue sample from the
XX tumour expresses a polypeptide comprising at least 14 continuous amino
XX acids of the protein encoded by this sequence. The method is useful for
XX determining the risk of metastasis of a female breast tumour and to
XX predict the response of the female breast tumour to hormone treatment.
XX pLIV1 expression is highly predictive of tumour oestrogen receptor
XX expression and lymph node involvement.
XX
SQ Sequence 2404 BP; 726 A; 411 C; 492 G; 775 T; 0 other:
Query Match 55.7%; Score 1263; DB 19; Length 2404;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1285; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 982 GGTTTTATAGCCATTTCCATCATCAGTTTCTGCTCTGCTGGGGTATCTTAGTGCT 1041
Db 1 GGTTTTATAGCCATTTCCATCATCAGTTTCTGCTCTGCTGGGGTATCTTAGTGCT 60
QY 1042 CTCATGAATCGGGTGTGTTTTTCCAAATTTCTCCTGAGTTTCTTGGCAGCTGGCGTGGG 1101
Db 61 CTCATGAATCGGGTGTGTTTTTCCAAATTTCTCCTGAGTTTCTTGGCAGCTGGCGTGGG 120
QY 1102 ACTTTGAGTGTGATGCTTTTTTACACCTTCTCCACATCTTCATGCAAGTCACCCAT 1161
Db 121 ACTTTGAGTGTGATGCTTTTTTACACCTTCTCCACATCTTCATGCAAGTCACCCAT 180
QY 1162 AGTCATAGCCATGAAGAACCCAGCAATGGAATGAAAGAGGAGCCACTTTTTCAGTCATCTG 1221
Db 181 AGTCATAGCCATGAAGAACCCAGCAATGGAATGAAAGAGGAGCCACTTTTTCAGTCATCTG 240
QY 1222 TCTTCTCAAAACATAGAGAAAGTGCCCTATTTTCATTTCCAGTGGAAGGTCTTAACAGCT 1281
Db 241 TCTTCTCAAAACATAGAGAAAGTGCCCTATTTTCATTTCCAGTGGAAGGTCTTAACAGCT 300
QY 1282 CTAGGAGGCTGTATTTTCATGTTTCTGTTGAACATCTCCTCACATTTGATCAACAATTT 1341
Db 301 CTAGGAGGCTGTATTTTCATGTTTCTGTTGAACATCTCCTCACATTTGATCAACAATTT 360
QY 1342 AAGATAAAGAGAAAAAGAAATCAGAAGAAACCTGAAATGATGATGATGATGATGATGAT 1401
Db 361 AAGATAAAGAGAAAAAGAAATCAGAAGAAACCTGAAATGATGATGATGATGATGATGAT 420
QY 1402 AAGCAGTTGTTCCAAAGTATGAATCTCAACTTTTCAACAATGAGGAGAAAGTAGATACAGAT 1461
Db 421 AAGCAGTTGTTCCAAAGTATGAATCTCAACTTTTCAACAATGAGGAGAAAGTAGATACAGAT 480
QY 1462 GATCGAACTGAAGCTATTTACGAGCAGACTCAACAAGCCCTCCCACTTTGATTTCTCAG 1521
Db 481 GATCGAACTGAAGCTATTTACGAGCAGACTCAACAAGCCCTCCCACTTTGATTTCTCAG 540
QY 1522 CAGCTGCAGTCTTTGGAGAGAGAGAGTTCATGATAGTCTCATGCTCATCCACAGGAAGTC 1581

```
Dh 541 CAGCCTGAGTCTTTGGAAGAGAGAGGTGATGATAGTCTATGCTCATCCACAGGAAGTC 600
Qy 1582 TACAATCAATATGTACCCAGAGGTGCAAGAATAAATGCCATTACATTTCCACGATACA 1641
Dh 601 TACAATCAATATGTACCCAGAGGTGCAAGAATAAATGCCATTACATTTCCACGATACA 660
Qy 1642 CTCGGCCAGTCAGAGATCTCATTTCCACCACCATCATGACTGACTGATGATATTTCCATCAT 1701
Dh 661 CTCGGCCAGTCAGAGATCTCATTTCCACCACCATCATGACTGACTGATGATATTTCCATCAT 720
Qy 1702 CACCACCACCAAAACACCATCTCCACAGTCACAGCCAGGCGTACTCTCGGAGGAGCTG 1761
Dh 721 CACCACCACCAAAACACCATCTCCACAGTCACAGCCAGGCGTACTCTCGGAGGAGCTG 780
Qy 1762 AAAGATGCGCGCGTCCGACCTTTGGCGTGGATGATGATGATGATGATGATGATGATGATGAT 1821
Dh 781 AAAGATGCGCGCGTCCGACCTTTGGCGTGGATGATGATGATGATGATGATGATGATGATGAT 840
Qy 1822 TTACGCGATGGCTAGCAATTTGGTGTCTTTTACTGAAGGCTTATCAAGTGTCTTAAAGT 1881
Dh 841 TTACGCGATGGCTAGCAATTTGGTGTCTTTTACTGAAGGCTTATCAAGTGTCTTAAAGT 900
Qy 1882 ACTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941
Dh 901 ACTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy 1942 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGATGATGAT 2001
Dh 961 CTAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGATGATGAT 1020
Qy 2002 GCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTCAATTATGCTGAAATGTTTCTATG 2061
Dh 1021 GCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTCAATTATGCTGAAATGTTTCTATG 1080
Qy 2062 TGGATATTTGCACTTACTGCTGCTTATTCATGATGATGATGATGATGATGATGATGATGATGAT 2121
Dh 1081 TGGATATTTGCACTTACTGCTGCTTATTCATGATGATGATGATGATGATGATGATGATGATGAT 1140
Qy 2122 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGATGATGATGATGATGATGATGATGATGAT 2181
Dh 1141 GAAATGCTGCACAAATGATGCTAGTGACCATGGATGATGATGATGATGATGATGATGATGATGAT 1200
Qy 2182 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATGATGATGATGATGATGATGATGATGATGATGAT 2241
Dh 1201 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATGATGATGATGATGATGATGATGATGATGATGAT 1259
Qy 2242 AAAATGCTGTTTGGTATAAATTTCTAG 2268
Dh 1260 AAAATGCTGTTTGGTATAAATTTCTAG 1285

RESUIR 8
AA33219
ID AAT33219 standard; cDNA; 1310 BP.
XX
AC AAT33219;
XX
AC XX
XX
DT 16-OCT-1996 (first entry)
XX
DE Oestrogen-regulated pLIV1 gene partial cDNA clone.
XX
KW Oestrogen-regulated gene; pLIV1; breast cancer; tumour; diagnosis;
KW market; metastasis; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS 1..1296
FT /*tag= a
XX
XX CA2132500-A.
PN
XX 21-MAR-1996.
PD
```

```
XX 20-SEP-1994; 94CA-2132500.
PF
XX 20-SEP-1994; 94CA-2132500.
PR
XX (UYWA-) UNIV WALES COLLEGE OF MEDICINE.
PA
PI Gee JM, Green CD, Manning DL, Nicholson RI;
PI WPI: 1996-268987/28.
DR P-PSDB; AAR98004.
XX
PT Oestrogen-regulated pLIV1 gene and deduced polypeptide - useful for
PT predicting the propensity for metastatic spread or the
PT responsiveness to endocrine treatment of breast tumour
XX
XX Claim 20; Page 12-14; 27pp; English.
XX
CC The oestrogen-regulated pLIV1 gene (partial cDNA given in AAT33219)
CC is a candidate for the hormonal regulation of tumour invasion of
CC breast cancer, a prognostic marker of metastatic spread, and a
CC suitable target for therapeutic intervention by anti-hormones and
CC antibody-directed methods. It is significantly associated with
CC oestrogen-dependent (ER+) disease. The pLIV1 coding sequence was
CC isolated from a series of oestrogen-regulated genes. A 3' extended
CC version (AAT33220) of the partial cDNA clone was also detd. The
CC polypeptide (AAR98004) encoded by the partial cDNA contains 3
CC immunogenic regions.
XX
SQ Sequence 1310 BP; 354 A; 273 C; 285 G; 398 T; 0 other;
Query Match 55.5%; Score 1259.8; DB 17; Length 1310;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1283; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
Qy 982 GGTGTTTATAGCCATTTCATCATCATCAGTTTCCTGCTGCTGGGGTGTATCTTAGTGCT 1041
Dh 1 GGTGTTTATAGCCATTTCATCATCATCAGTTTCCTGCTGCTGGGGTGTATCTTAGTGCT 60
Qy 1042 CTCATGAATCGGGTGTGTTTTCAAATTTCTCGTAGTTTCTTGTGGCACTGGCGGTGGG 1101
Dh 61 CTCATGAATCGGGTGTGTTTTCAAATTTCTCGTAGTTTCTTGTGGCACTGGCGGTGGG 120
Qy 1102 ACTTTGAGTGGTGATGCTTTTTCACACCTTCTCCACATTTCTCATGCAAGTCACCACT 1161
Dh 121 ACTTTGAGTGGTGATGCTTTTTCACACCTTCTTCCACATTTCTCATGCAAGTCACCACT 180
Qy 1162 AGTCATAGCCATGAAGAACCCAGCAATGGAATGAAAGAGGACCACTTTTCAGTCACTG 1221
Dh 181 AGTCATAGCCATGAAGAACCCAGCAATGGAATGGAATGAAAGAGGACCACTTTTCAGTCACTG 240
Qy 1222 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTGTATTCACGTGGGAGGCTTAACAGCT 1281
Dh 241 TCTTCTCAAAACATAGAGAAAGTGCCTATTTTGTATTCACGTGGGAGGCTTAACAGCT 300
Qy 1282 CTAGGAGCGCTGATTTTCATGCTTTTCTTGTGACATGTCCTCACATTTGATCAAACTTT 1341
Dh 301 CTAGGAGCGCTGATTTTTCATGCTTTTCTTGTGACATGTCCTCACATTTGATCAAACTTT 360
Qy 1342 AAAGATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1401
Dh 361 AAAGATAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy 1402 AAGCAGTTGTCAGATGATGAATCTCAACTTTTCAACAAATGAGGAGAGAGAGAGAGAGAG 1461
Dh 421 AAGCAGTTGTCAGATGATGAATCTCAACTTTTCAACAAATGAGGAGAGAGAGAGAGAGAG 480
Qy 1462 GATCGAAGTGAAGCTATTTACGAGCAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAG 1521
Dh 481 GATCGAAGTGAAGCTATTTACGAGCAGACTCACAAGAGCCCTCCCACTTTGATTTCTCAG 540
Qy 1522 CAGCTGCGAGTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1581
Dh 1522 CAGCTGCGAGTCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
```

Ub	541	CAUCCTGCAGTCCTTTGGAGAAGAAAGAGGTCATGATACCTTCATGCTCCTCAGGAAATCTC	600
Qy	1582	TACAAATCAATATATACCCAGAGGGTGCAGAATAAATGCCATTTCACATTTCTACGATACA	1641
Dh	601	TACAATCAATATCTACCCAGAGGGTGCAGAATAAATGCCATTTCACATTTCTACGATACA	660
Qy	1642	CTCGGCAGTCAGACGATCTCATTTCCACACCAATCATGACTACCAATCATATTTCTGCATCAT	1701
Dh	661	CTCGGCAGTCAGACGATCTCATTTCCACACCAATCATGACTACCAATCATATTTCTGCATCAT	720
Qy	1702	CACCACCAACCAACCACTCTCACAGTCACAGCCAGGCGTACTCTCGGAGAGACCTG	1761
Dh	721	CACCACCAACCAACCACTCTCACAGTCACAGCCAGGCGTACTCTCGGAGAGAGCTG	780
Qy	1762	AAAGATCCCGGCTCGCCACTTTTGGCCTGGATGTGATAATAGGGTCATGGCTGCGACAAT	1821
Dh	781	AAAGATCCCGGCTCGCCACTTTTGGCCTGGATGTGATAATAGGGTCATGGCTGCGACAAT	840
Qy	1822	TTCAGCGATGGGCTTACCAATTTGGTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGT	1881
Dh	841	TTCAGCGATGGGCTTACCAATTTGGTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAGT	900
Qy	1882	ACTTCTCTTGGCTGTGTCTGTCTCATGAGTTGGCTCATGAATAGGTGACTTTTGCCTGTTCTA	1941
Dh	901	ACTTCTCTTGGCTGTGTGTCTGTCTCATGAGTTGGCTCATGAATAGGTGACTTTTGCCTGTTCTA	960
Qy	1942	CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCCCTTTTAAATGCAATCTCGACCATGCTG	2001
Dh	961	CTAAGGCTGGCATGACCGTTAAGCAGGCTGTCCCTTTTAAATGCAATCTCGACCATGCTG	1020
Qy	2002	GCCTATCTTTGGAATGGCAACAGGAAATTTTCATTGGCTCATTTAGCTGAAATGTTTCTATG	2061
Dh	1021	GCCTATCTTTGGAATGGCAACAGGAAATTTTCATTGGCTCATTTAGCTGAAATGTTTCTATG	1080
Qy	2062	TGGATATTTGCACTTACTGCTGGCTTATTCATGTATGTTGCTCTGGTTGATATGGTACCT	2121
Dh	1081	TGGATATTTGCACTTACTGCTGGCTTATTCATGTATGTTGCTCTGGTTGATATGGTACCT	1140
Qy	2122	GAATGCTGCACAAATGCTAGTGACCATGGATGATGCCGCTGGGGGTATTTCTTTTAA	2181
Dh	1141	GAATGCTGCACAAATGCTAGTGACCATGGATGATGCCGCTGGGGGTATTTCTTTTAA	1200
Qy	2182	CAGAATGCTGGGATGCTTTTGGGTTTGGGAATATGTTTACTTATTTTCCATATTTTGAACAT	2241
Dh	1201	CAGAATGCTGGGATGCTTTTGGGTTTGGGAATATGTTTACTTATTTTCCATATTTTGAACAT	2260
Qy	2242	AAATCGTGTTCGGTATAAATTTCTAG	2268
Dh	1260	AAATCGCTG-TTTCGTATAAATTTCTAG	1285

RESULT 9	
AAT33220	
11)	AAT33220 standard; cDNA; 2404 BP.
XX	
XX	
AC	AAT33220;
XX	
XX	16-OCT-1996 (first entry)
DT	
XX	
DE	Oestrogen-regulated pL1V1 gene partial 3' extended cDNA clone.
XX	
XX	Oestrogen-regulated gene; pL1V1; breast cancer; tumour; diagnosis;
KW	marker; metastasis; ss.
KW	

OS	XX	Key	Location/Qualifiers
homo sapiens.	XX	Key	1..1296
	XX	CDS	/*tag= a
	XX	3'UTR	1297..2404
	XX		/*tag= b
	XX	misc_difference	2201..2260
	XX		/*tag= c

FT	/note= "bases 2201-2260 are identical to bases
FT	2141-2200"
PN	CA2132500-A.
XX	21-MAR-1996.
PD	20-SEP-1994; 94CA-2132500.
XX	20-SEP-1994; 94CA-2132500.
PF	(UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX	Gea JM, Green CD, Manning DL, Nicholson RI;
PI	WPI; 1996-268987/28.
XX	P-PSDB; AAR98004.
DR	Oestrogen-regulated p11v1 gene and deduced polypeptide - useful for
DR	predicting the propensity for metastatic spread or the
XX	responsiveness to endocrine treatment of breast tumour
PT	Disclosure; Page 16-17; 27pp; English.
PT	A cDNA clone (AAT33219) corresponding to a portion of the oestrogen-
XX	regulated human p11v1 gene was extended (AAT33220) using a primer
CC	directed cloning strategy. p11v1 is a candidate for the hormonal
CC	regulation of tumour invasion of breast cancer, a prognostic marker
CC	of metastatic spread, and a suitable target for therapeutic
CC	intervention by antihormones and antibody-directed methods. It is
CC	significantly associated with oestrogen-dependent (ER+) disease.
CC	The polypeptide (AAR98004) encoded by the cDNA contains 3 immunogenic
CC	regions.
XX	Sequence 2404 BP: 726 A; 413 C; 490 G; 775 T; 0 other:
XX	

Query Match	55.5%	Score 1259.8	DB 17	Length 2404
Best Local Similarity	99.7%	Pred. No. 0		
Matches 1283	Conservative	0	Mismatches	2
			Indels	Gaps
Qy 982	GGTTTATTATAGCAATTCATCATCAGATTCTCTGTCTCTGCTGGGGTTATCTTTAGTGCCT	1041		
Db 1	GGTTTATTATAGCCATTTCCATCATCAGTTTCTGTCTCTGCTGGGGTTATCTTTAGTGCCT	60		
Qy 1042	CTCATGAATCGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGTGGCAGCTGGCCGTTGGG	1101		
Db 61	CTCATGAATCGGGTGTTTTTCAAAATTTCTCTGAGTTTCCCTGTGGCAGCTGGCCGTTGGG	120		
Qy 1102	ACTTTGAGTGGTGATGCTTTTTTTACACCTTCTCCACATTCCTCATGCAAGTCACCAACAT	1161		
Db 121	ACTTTGAGTGGTGATGCTTTTTTTACACCTTCTCCACATTCCTCATGCAAGTCACCAACAT	180		
Qy 1162	AGTCATAGCCCATGAAGAACAGCAATGGAATGAAAGAGGAGCACCACATTTTCAGTCATCTG	1221		
Db 181	AGTCATAGCCCATGAAGAACAGCAATGGAATGAAAGAGGAGCACCACATTTTCAGTCATCTG	240		
Qy 1222	TCTTCTCAAACATAGAGAAAGTGCCCTATTTTCATTTCCAGCTGCGAAGGGTCTTAACAGCT	1281		
Db 241	TCTTCTCAAACATAGAGAAAGTGCCCTATTTTCATTTCCAGCTGCGAAGGGTCTTAACAGCT	300		
Qy 1282	CTAGAGGCCGTGTAATTCATGTTTCTTTGTTGAACATGTCCTCACATTTGATCAAACAATTT	1341		
Db 301	CTAGAGGCCGTGTAATTCATGTTTCTTTGTTGAACATGTCCTCACATTTGATCAAACAATTT	360		
Qy 1342	AAAGATAGAAGAAAAAGAAATCAGAAGAACCTGAAATGATGATGATGGAGATTAAAG	1401		
Db 361	AAAGATAGAAGAAAAAGAAATCAGAAGAACCTGAAATGATGATGATGGAGATTAAAG	420		
Qy 1402	AAGCAGTTGTGCCAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT	1461		
Db 421	AAGCAGTTGTGCCAAGTATGAATCTCAACTTTCAACAATGAGGAGAAAGTAGATACAGAT	480		
Qy 1462	GATCGAACTGAAGGCTATTTACGAGCAGACTCAAGAGGCCCTCCCACTTTGATTTCTCAG	1521		

```
db 481 GATCGAACTCAAGGCTATTTACAGCAGACTCACAGAGCCCTCCACATTTGATTTCTCAG 540
QY 1522 CAGCCTGCGAGTCTTGGAAGAAGAGGTCATGATGCTCATGCTCATCAACAGGAAGTC 1581
db 541 CAGCCTGAGTCTTGGAAGAAGAGGTCATGATGCTCATGCTCATCAACAGGAAGTC 600
QY 1582 TACAATGAATATGTATCCAGAGAGGGTGCAAGAATAAATGCCATTTCACATGATACA 1641
db 601 TACAATGAATATGTATCCAGAGAGGGTGCAAGAATAAATGCCATTTCACATGATACA 660
QY 1642 CTGCGGCGAGTCAGACATCTTACACACCATCATGATGCTCATGCTCATCAATTTCCATCAT 1701
db 661 CTGCGGCGAGTCAGACATCTTACACACCATCATGATGCTCATGCTCATCAATTTCCATCAT 720
QY 1702 CACCACCAACCAACCACTCTACAGTCACAGCAGCGTACTCTCGGAGAGGCTG 1761
db 721 CACCACCAACCAACCACTCTACAGTCACAGCAGCGTACTCTCGGAGAGGCTG 780
QY 1762 AAAGATGCGCGGCTGCGCCACTTTGGCCCTGGATGGTGATAGGTCATGCGCTGCACAA 1821
db 781 AAAGATGCGCGGCTGCGCCACTTTGGCCCTGGATGGTGATAGGTCATGCGCTGCACAA 840
QY 1822 TTTCAGCGATGCGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTCTTTAAGT 1881
db 841 TTTCAGCGATGCGCTAGCAATTTGGTCTGCTTTTACTGAAGGCTTATCAAGTCTTTAAGT 900
QY 1882 ACTTCGTTGCTGCTGCTGCTCATGATGCTGCTCATGATGCTGCTGCTGCTGCTGCTGCT 1941
db 901 ACTTCGTTGCTGCTGCTGCTCATGATGCTGCTCATGATGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 1942 CTAAAGGCTGGCATGACCTTAAGCAGGCTGCTCTTTTATAATGCAATTTGTCAGCCATGCTG 2001
db 961 CTAAAGGCTGGCATGACCTTAAGCAGGCTGCTCTTTTATAATGCAATTTGTCAGCCATGCTG 1020
QY 2002 GCGTATCTTGGAAATGCAACAGCAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2061
db 1021 GCGTATCTTGGAAATGCAACAGCAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 2062 TGCGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2121
db 1081 TGCGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 2122 GAAATGCTGCACATGATGCTAGTGACCATGATGCTAGCGGCTGCTGCTGCTGCTGCTGCTGCT 2181
db 1141 GAAATGCTGCACATGATGCTAGTGACCATGATGCTAGCGGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 2182 CAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2241
db 1201 CAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
QY 2242 AAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2268
db 1260 AAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1285

RESULT 10
ID AAF16156
XX AAF16156 standard; cDNA; 1193 BP.
AC AAF16156;
XX
XX 13-MAR-2001 (first entry)
DE human prostate cancer antigen nucleotide sequence SEQ ID NO:591.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioprotective; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neutral; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
```

```
OS Homo sapiens.
XX WO200005174-A1.
PN 21-SEP-2000.
PD
XX 08-MAR-2000; 2000WO-US05988.
PF 12-MAR-1999; 99US-0124270.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM;
XX WPI: 2000-587513/55.
DR P-PSDB: AAB56953.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX Claim 1: Page 1045-1046; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytoskeletal,
XX cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
XX nephrotropic, antiinfective, gynaecological and antibacterial activities,
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
SQ Sequence 1193 BP; 393 A; 259 C; 235 G; 302 T; 4 other;

Query Match 50.3%; Score 1141.4; DB 21; Length 1193;
Best Local Similarity 99.7%; Pred. No. 1.2e-301;
Matches 1139; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 539 ACAGTGTAGTCTAGTCAAGTGAACCTCAACTGTGTACACACTGTCTCTGAAGGAACCTC 598
Db 1 ACAGTGTAGTCTAGTCAAGTGAACCTCAACTGTGTACACACTGTCTCTGAAGGAACCTC 60
QY 599 ACTTTCTAGACACAATAGAGACTCCAAAGACTTGGAAACCTTCTCCCAAGAGACTTAAGCA 658
Db 61 ACTTTCTAGACACAATAGAGACTCCAAAGACTTGGAAACCTTCTCCCAAGAGACTTAAGCA 120
QY 659 GCTCCACTCCACCAGTGTACATCAAGAGCGGGGTGAGCGGCTGGCTGGTAGGAAAA 718
Db 121 GCTCCACTCCACCAGTGTACATCAAGAGCGGGGTGAGCGGCTGGCTGGTAGGAAAA 180
QY 719 CAATGAATCTGTGAGTCAGCCCGGAGAGGCTTTATGTATTCCAGAAACACAAATGAAA 778
Db 181 CAATGAATCTGTGAGTCAGCCCGGAGAGGCTTTATGTATTCCAGAAACACAAATGAAA 240
QY 779 ATCCCTCAGGAGTGTTCATGATCAAGCTTACTGACATCTCATGGCATGGGATCCAGG 838
Db 241 ATCCCTCAGGAGTGTTCATGATCAAGCTTACTGACATCTCATGGCATGGGATCCAGG 300
QY 839 TTCGCGTGAATGCAACAGAGTTCAACTATCTCTCTCCAGGCATCATCAACCAAAATGATG 898
Db 301 TTCGCGTGAATGCAACAGAGTTCAACTATCTCTCTCCAGGCATCATCAACCAAAATGATG 360
QY 899 CTAGATCTTGTCTGATTCATACAGTCAAGAGAGGCTGAAATCCCTCCAAAGACCTATT 958
Db 361 CTAGATCTTGTCTGATTCATACAGTCAAGAGAGGCTGAAATCCCTCCAAAGACCTATT 420
```



```
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
PT
```

```
PT metastasis -
XX
PS Disclosure; SEQ ID NO 34631; 307lpp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 20778 BP; 5829 A; 4050 C; 4125 G; 6774 T; 0 other;
Query Match 31.7%; Score 787.4; DB 22; Length 20778;
Best Local Similarity 99.9%; Pred. No. 4.1e-204;
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGGCGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTTGGCCCTCTCTGTACAAATCCC 60
Db 2297 ATGGCGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTTGGCCCTCTCTGTACAAATCCC 2356
Qy 61 CTTTCATGAACATAAAGCAGCTGCTTCCCCAGACCACTGAGAAAATAGTCCCAATTGG 120
Db 2357 CTTTCATGAACATAAAGCAGCTGCTTCCCCAGACCACTGAGAAAATAGTCCCAATTGG 2416
Qy 121 GAATCTGGCATTAAATGTTGACTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTT 180
Db 2417 GAATCTGGCATTAAATGTTGACTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTT 2476
Qy 181 TTCTACCGCTATGGAGAAAATAATCTTTGTGCTTGAAGGTTTCAGAAAATTTACTTCAA 240
Db 2477 TTCTACCGCTATGGAGAAAATAATCTTTGTGCTTGAAGGTTTCAGAAAATTTACTTCAA 2536
Qy 241 AATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACACGACCATCACTCA 300
Db 2537 AATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACACGACCATCACTCA 2596
Qy 301 GACCACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCATC 360
Db 2597 GACCACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCATC 2656
Qy 361 GAGCATCACTCTGACCATGATCATCACTTCCACCATATCATGCTTCTGGTAAAAAT 420
Db 2657 GAGCATCACTCTGACCATGATCATCACTTCCACCATATCATGCTTCTGGTAAAAAT 2716
Qy 421 AAGCGAAAAGCTCTTTGCCCGAGACCATGACTCAGATAGTTCAGGTAAAGATCCCTAGAAAC 480
Db 2717 AAGCGAAAAGCTCTTTGCCCGAGACCATGACTCAGATAGTTCAGGTAAAGATCCCTAGAAAC 2776
Qy 481 AGCCAGGGGAAAGGAGCTCACCAGCAGAACATGCCAGTGGTAGAAGGAATGTCAAGGAC 540
Db 2777 AGCCAGGGGAAAGGAGCTCACCAGCAGAACATGCCAGTGGTAGAAGGAATGTCAAGGAC 2836
Qy 541 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACACACTGTCTCTGAAGGAACACTCAC 600
Db 2837 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACACACTGTCTCTGAAGGAACACTCAC 2896
Qy 601 TTTCTAGACACAATAGACACTCCCAAGACCTGGAACCTTTCCCAAGAGATGTGAAGCAGC 660
Db 2897 TTTCTAGACACAATAGACACTCCCAAGACCTGGAACCTTTCCCAAGAGATGTGAAGCAGC 2956
Qy 661 TCCACTCCACCCAGTGTGCATCATCAAAAGAGCCGGGTGAGCCGGCTGGCTGGTAGGAAAACA 720
```


Db	721	AGAGGAAGANGTCATGATAGCCCATGACACACCCACAGAAAGTCTTACAAATGATATGCTG	780
QY	1596	ACCCAGAGGGTCCAGAAATAATGCCATTTCCATTTCCACGATACACTCGGCCAATCAGA	1655
DB	781	CCACAGGGGCTCCAGAGAAAGTGCCATTCCACACTTTACGATACGCTGGGCACTCCGA	840
QY	1656	CVATCTCATTTACCAACCACATCATGACTACCATCATATTTCTTCATCATCACCAC	1707
DB	841	CUACCTT-ATNCAACCACCATCAGACTACCATCATCATTCGNAACCAACACCTTC	891
RESULT 13			
AA013487	AA013487 standard; cDNA; 551 BP.		
XX	XX	AA013487:	
XX	XX	06-NOV-2001 (first entry)	
XX	XX	Human LIV-1-164647 cDNA fragment (nucleotides 1690-2240).	
DE	DE	Human; LIV-1-164647; cytostatic; estrogen-inducible gene; tumour;	
XX	KW	cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;	
KW	KW	salivary gland; carcinoma; drug screening; therapy; ss.	
XX	XX	Homo sapiens.	
OS	OS	W0200155178-A2.	
XX	XX	02-AUG-2001.	
XX	XX	25-JAN-2001; 2001WO-0502622.	
XX	XX	25-JAN-2000; 2000US-0177951.	
XX	PK	10-APR-2000; 2000US-0195761.	
XX	XX	(CETH) GENENTECH INC.	
XX	PA	Giddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;	
XX	PI	WPI; 2001-502628/55.	
XX	XX	New LIV-1 RNA and amino acids, useful for treating cancer or inhibiting	
XX	PT	the proliferation of tumor cells in mammals, e.g. breast, lung,	
XX	PT	prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma	
XX	PT	Example 5; Page 76-77; 150pp; English.	
XX	XX	The present sequence is human estrogen-inducible LIV-1-164647 cDNA	
CC	CC	fragment (nucleotides 1690-2240). LIV-1 is overexpressed in tumour	
CC	CC	tissues such as prostate, colon, lung and breast. The LIV-1 DNA and	
CC	CC	polypeptide, and the monoclonal antibody are useful for treating cancer	
CC	CC	and inhibiting the proliferation of tumour cells in mammals, e.g. breast,	
CC	CC	lung, prostate, colon, ovary, uterus, kidney, gastric or salivary	
CC	CC	carcinoma, or other tumour cell types expressing the LIV-1-164647	
CC	CC	protein. In particular, the mammal is a human. the LIV-1 DNA and	
XX	XX	polypeptide may also be used in screening assays for drug candidates.	
XX	SQ	Sequence 551 BP; 128 A; 130 C; 127 G; 166 T; 0 other;	

Qy	1738	CACGCGTACTCTCGGAGGAGCTGAAGATGATCCGCGCTGCACACTTTGGCCTTGGATGGTG	1799
Db	121	CACGCGTACTCTCGGAGGAGCTGAAGATGATCCGCGCTGCACACTTTGGCCTTGGATGGTG	180
Qy	1798	ATAATGGGTGATGGCTGCACAAATTTTCAGCGATAGCCCTAGCAATTTGGTGTCTTTTACT	1857
Db	181	ATAATGGGTGATGGCTGCACAAATTTTCAGCGATAGCCCTAGCAATTTGGTGTCTTTTACT	240
Qy	1858	GAAGCGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGCTGTGTCTTGTTCATGAGTTGCCCTCAT	1917
Db	241	GAAGCGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGCTGTGTCTTGTTCATGAGTTGCCCTCAT	300
Qy	1918	GAATTAGGTGACCTTGTCTGTCTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGTCTT	1977
Db	301	GAATTAGGTGACCTTGTCTGTCTACTAAAGGCTGACATGACCGTTAAGCAGGCTGTCTT	360
Qy	1978	TATAATGCATTGTTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTGGT	2037
Db	361	TATAATGCATTGTTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTGGT	421
Qy	2038	CATTATGCTGAAAATGTTTCTATGTGGATATTTGGCACTTACTTGTGGCTTATTTCATGTAT	2097
Db	421	CATTATGCTGAAAATGTTTCTATGTGGATATTTGGCACTTACTTGTGGCTTATTTCATGTAT	480
Qy	2098	GTTCGCTCGTTGATATGGTACTGAAATGCTGCAATGATGCTAGTGACCATGGATGT	2157
Db	481	GTTCGCTCGTTGATATGGTACTGAAATGCTGCAATGATGCTAGTGACCATGGATGT	540
Qy	2158	AGCGCGTGGGG 2168	
Db	541	AGCGCGTGGGG 551	
RESULT 14			
AAZ33622			
ID	AAZ33622 standard; cDNA; 1597 BP.		
AC	AAZ33622;		
XX			
XX	08-DEC-1999 (first entry)		
XX	Human breast tumour-associated EST 12.		
DE			
XX	Expressed sequence tag; EST; human; breast; cancer; gene therapy;		
KW	treatment; tumour; cylostatic; medicament; ss.		
KW			
XX	Homo sapiens.		
OS			
XX	DE19813839-A1.		
XX	23-SEP-1999.		
XX			
XX	20-MAR-1998; 98DE-1013839.		
XX			
XX	20-MAR-1998; 98DE-1013839.		
XX	(META-) METAGEN GES GENOMFORSCHUNG MBH.		
PA			
XX	Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosentahl A;		
XX	WPI; 1999-528981/45.		
XX			
XX	Human nucleic acid sequences and protein products from tumor breast		
PT	tissue, useful for breast cancer therapy -		
XX			
XX	Claim 3; 93; 188pp; German.		
PS			
XX			
CC	This invention describes novel human nucleic acid sequences from tumor		
CC	breast tissue which have cytostatic activity. The nucleic acid sequences		
CC	can be used to produce and isolate full-length gene sequences. They can		
CC	be used to express proteins, which can be used as tools to find an		
CC	activity against breast cancer. The sequences can be used in sense or		
CC	antisense form. They are especially useful for medicaments for gene		

Db	1818	TACCATCATATTCTCCATCATACACCAACCAACCAACCATCCTCACAGTCACACCCAG	1877
Qy	1741	CGCTACTCTCGGGAGAGCTGAAAGATGCCGGCGTCGCCACTTTGGCTCGGATGGTGATA	1800
Db	1878	CGCTACTCTCGGAGGAGCTGAAGATGCCGGCGTCGCCACTTTGGCTCGGATGGTGATA	1937
Qy	1801	ATGGGTGATGGCCTCACAATTTCAGCGATGGCTAGCAATGGTGCTGCTTTACTGAA	1860
Db	1938	ATGGGTGATGGCCTGCACAATTTCAGCGATGGCTAGCAATGGTGCTGCTTTACTGAA	1997
Qy	1861	GGCTTATCAAGTGGTTTAAGTACTCTCTGGCTGCTGTTCTGTCTCATGAGTGGCTCATGAA	1920
Db	1998	GGCTTATCAAGTGGTTTAAGTACTCTCTGGCTGCTGTTCTGTCTCATGAGTGGCTCATGAA	2057
Qy	1921	TTAGGTGACTTTGGCTGTTCTACTAAAGGCTGGCATGACCTTAAAGCAGGCTGCTCTTTAT	1980
Db	2058	TTAGGTGACTTTGGCTGTTCTACTAAAGGCTGGCATGACCTTAAAGCAGGCTGCTCTTTAT	2117
Qy	1981	AATGCATTCTCAGCCATGCTGGCGTATCTTGGAAATGCCAACAGAGAAATTTTCATGGTCAT	2040
Db	2118	AATGCATTCTCAGCCATGCTGGCGTATCTTGGAAATGCCAACAGAGAAATTTTCATGGTCAT	2177
Qy	2041	TATGCTGAAAAATGTTTCTATGTGGATATTTCACATTACTGCTGGCTATTTCATGTATGTT	2100
Db	2178	TATGCTGAAAAATGTTTCTATGTGGATATTTCACATTACTGCTGGCTATTTCATGTATGTT	2237
Qy	2101	GCTCTGGTTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGAGC	2160
Db	2238	GCTCTGGTTGATATGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGAGC	2297
Qy	2161	CGCTGGGGCTATTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2220
Db	2298	CGCTGGGGCTATTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2357
Qy	2221	CTTATTTCCATATTGAAACATAAAATCGTGTTCGTATAAAATTTCTAG	2268
Db	2358	CTTATTTCCATATTGAAACATAAAATCGTGTTCGTATAAAATTTCTAG	2405

RESULT 2			
AX207207			
LOCUS	AX207207	2776 bp	DNA
DEFINITION	Sequence 3 from Patent WO0155178.		linear
			PAT 30-AUG-2001

ACCESSION	AX207207
VERSION	AX207207.1
KEYWORDS	GI:15394961
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 2776)
TITLE	Goddard, A., Gurney, A.L., Smith, V., Hongo, J.A. and de Sauvage, F.
JOURNAL	Compositions and methods for treatment of cancer Patent: WO 0155178-A 3 02-AUG-2001; GENENTECH, INC. (US)
FEATURES	Location/Qualifiers
source	1..2776

```

/organism="Homo sapiens"
/db_xref="taxon:9606"

823 a      583 c      592 g      778 t

BASE COUNT
ORIGIN

```

Query Match 99.9%; Score 2264.8; DB 6; Length 2776;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2266; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCAGGAAGTTATCTGTAATCTTGATCTCTGACCTTGGCTCTCTCTGTCACAAATCCC 60
|||||
73 ATGGCAGGAGCTTATCTGTAATCTTGATCTGACCTTGGCTCTCTCTGTCACAAATCCC 132
Db
Qy 61 CTTTATGACTAAAGCAGCTGCTTTCCCCAGCACCACCTGAGAAAAATTAGTCGGAAATGG 120
|||||

Db	133	CTTCATGAACATAAAGACAGCTGTTTCCGCCCCAGACCACCTAGAGAAATTAGTCCGGAATTGG	199
Qy	121	GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGCGAATATCATCTACAACAGCTT	180
Db	193	GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACACGCGAATATCATCTACAACAGCTT	252
Qy	181	TTCTACCGCTATGAGAGAAATAATCTTTGTCTCAGTTGAAGGGTTTCAGAAAAATTACTTCAA	240
Db	253	TTCTACCGCTATGAGAGAAATAATCTTTGTCTCAGTTGAAGGGTTTCAGAAAAATTACTTCAA	312
Qy	241	AATATAGGCATAGATTAAGATTTAAAGAAATCCATATACACCATGACACGACCATCACTCA	300
Db	313	AATATAGGCATAGATTAAGATTTAAAGAAATCCATATACACCATGACACGACCATCACTCA	372
Qy	301	GACCACGACATCAGACCATGAGCGTCTACTCAGACCATGAGCATCAGACCATCACTCAGACCA	360
Db	373	GACCACGACATCAGACCATGAGCGTCTACTCAGACCATGAGCATCAGACCATCACTCAGACCA	432
Qy	361	GAGCATCACTCTGACCATGATCATCTCTGCTCAACATAATCATGCTGCTTCTGGTAAAAAT	420
Db	433	GAGCATCACTCTGACCATGATCATCTCTGCTCAACATAATCATGCTGCTTCTGGTAAAAAT	492
Qy	421	AAGGAAAGCTCTTTGCCGACGACATGACTCAGATAGTTTCAGGTAAAGATCCTTGAAGAC	480
Db	493	AAGGAAAGCTCTTTGCCGACGACATGACTCAGATAGTTTCAGGTAAAGATCCTTGAAGAC	552
Qy	481	AGCAGGGGAAAGAGCTCAGCAGCAGAACATGCCAGTGGTAGAGGAATGTCAGGAC	540
Db	553	AGCAGGGGAAAGAGCTCAGCAGCAGAACATGCCAGTGGTAGAGGAATGTCAGGAC	612
Qy	541	AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGGAACCTCAC	600
Db	613	AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGGAACCTCAC	672
Qy	601	TTTCTAGAGACAATAGAGACTCCAAGACCTGGAAAACTCTTCCCCAAAGATGTAAAGCAG	660
Db	673	TTTCTAGAGACAATAGAGACTCCAAGACCTGGAAAACTCTTCCCCAAAGATGTAAAGCAG	732
Qy	661	TCACATCCACCCAGTGTACATCAAGAGCCGGTGCAGCCGGCTGCTGCTAGGAAACCA	720
Db	733	TCACATCCACCCAGTGTACATCAAGAGCCGGTGCAGCCGGCTGCTGCTAGGAAACCA	792
Qy	721	AATGAATCTGTAGTGAGCCCCGAAAAGGCTTTATGTATTCAGAAACACAAATGAAAT	780
Db	793	AATGAATCTGTAGTGAGCCCCGAAAAGGCTTTATGTATTCAGAAACACAAATGAAAT	852
Qy	781	CCTCAGGAGTGTTCATGATGCATCAAGCTACTGACATCTCATGGCAGTGGGCATCCAGGTT	840
Db	853	CCTCAGGAGTGTTCATGATGCATCAAGCTACTGACATCTCATGGCAGTGGGCATCCAGGTT	912
Qy	841	CCGCTGAATCAACAGAGTTCAACTATCTCTGCCAGCCATCATCAACCAAAATTCATGCT	900
Db	913	CCGCTGAATCAACAGAGTTCAACTATCTCTGCCAGCCATCATCAACCAAAATTCATGCT	972
Qy	901	AGATCTTGTCTGATTCATCAAGTGAAGAAAGGCTGAAATCCCTCCAAGACCTATTCA	960
Db	973	AGATCTTGTCTGATTCATCAAGTGAAGAAAGGCTGAAATCCCTCCAAGACCTATTCA	1032
Qy	961	TTACAAATAGCCGTGGTGGTGGTTTTATAGCCATTTCCATCATCAGTTTCTGCTCTG	1020
Db	1033	TTACAAATAGCCGTGGTGGTGGTTTTATAGCCATTTCCATCATCAGTTTCTGCTCTG	1092
Qy	1021	CTGGGGTTATCTTAGTGCCCTCATGAATCCGGGTGTTTTTCAAAATTTCTCTGAGTTTC	1080
Db	1093	CTGGGGTTATCTTAGTGCCCTCATGAATCCGGGTGTTTTTCAAAATTTCTCTGAGTTTC	1152
Qy	1081	CTTGTGGCACTGGCCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT	1140
Db	1153	CTTGTGGCACTGGCCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT	1212
Qy	1141	TCTCATGAAGTCAACCACTATAGTCTATGCCATGAGAAACCAAGCAATGAAATGAAAGA	1200
Db	1213	TCTCATGAAGTCAACCACTATAGTCTATGCCATGAGAAACCAAGCAATGAAATGAAAGA	1272

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:28:40 : Search time 5666 Seconds
(without alignments)
11649.341 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcgaggaagtattatctgt.....tggttcgtataaaattttctag 2268

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2268	100.0	3461	6	AX465588 Sequence
2	2264.8	99.9	2776	6	AX207207 Sequence
3	2222	98.0	2744	9	HS041060 Homo sapien
4	2198	96.9	3461	6	AX207205 Sequence
5	1616	71.3	3287	10	AB071697 Mus muscu
6	1263	55.7	1310	6	I76891 Sequence 1
7	1263	55.7	2404	6	I76892 Sequence 3
8	1151	50.7	2236	9	BC008317 Homo sapi
9	789	34.8	160170	2	AP001158 Homo sapi
10	787.4	34.7	221941	9	AC091060 Homo sapi
11	787.4	34.7	224788	2	AP001905 Homo sapi
12	619.6	27.3	899	6	AX465590 Sequence
13	549.4	24.2	551	6	AX207216 Sequence
14	509.4	22.5	1597	6	AX017261 Sequence
15	421.4	18.6	193437	2	AC126667 Rattus no
16	319.2	14.1	337	6	AX393861 Sequence
17	319	14.1	5231	9	AB033091 Homo sapi
18	319	14.1	5322	6	AX405756 Sequence
19	273.4	12.1	61103	2	AC131276 Homo sapi
20	191.2	8.4	193437	2	AC126667 Rattus no
21	144	6.3	52684	2	AC014796 Drosophil
22	144	6.3	174712	3	AC010041 Drosophil
23	144	6.3	195148	3	AC010019 Drosophil
24	144	6.3	282267	3	AE003555 Drosophil
25	129.2	5.7	2174	10	BC021530 Mus muscu
26	128.4	5.7	2015	9	AB051127 Macaca fa
27	127.6	5.6	4573	6	AX409002 Sequence
28	127.6	5.6	4573	9	HMORFKG1P Drosophil
29	126	5.6	4610	6	AX078294 Sequence
30	126	5.6	4698	6	AX281784 Sequence
31	123.4	5.4	1383	6	AX118897 Sequence
32	123.4	5.4	1803	9	AF193052 Homo sapi
33	123.4	5.4	2604	9	AC027652 Homo sapi
34	123.4	5.4	3169	9	BC012125 Homo sapi
35	123.4	5.4	3246	9	AB040120 Homo sapi
36	122.6	5.4	1895	9	BC027884 Homo sapi
37	119.2	5.3	1755	9	BC015770 Homo sapi
38	119.2	5.3	55577	2	AC014109 Drosophil
39	119.2	5.3	170071	3	AC093497 Drosophil
40	119.2	5.3	194979	3	AC091209 Drosophil
41	119.2	5.3	295377	3	AE003532 Drosophil
42	118	5.2	150	9	HUMDSFO4M5 Human HapG2
43	116.2	5.1	1931	10	BC028990 Mus muscu
44	115.8	5.1	3060	10	BC006731 Mus muscu
45	110.6	4.9	1888	6	AX061633 Sequence

ALIGNMENTS

RESULT 1	AX465588	Sequence 1	3461 bp	DNA	linear	PAT 16-JUL-2002
LOCUS	AX465588	Sequence 1 from Patent WO0216939.				
DEFINITION	AX465588					
ACCESSION	AX465588.1	GI:21899891				
VERSION						
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1					
AUTHORS	Mack,D., Gish,K.C. and Wilson,K.E.					
TITLE	Methods of diagnosis of cancer and screening for cancer modulators					
JOURNAL	Patent: WO 0216939-A 1 28-FEB-2002;					